ASAP: A Hybrid Computer Platform Using Machine Learning and Solar Imaging for Automated Prediction of Significant Solar Flares

T. Colak[1] and R. Qahwaji [2]

[1] University of Bradford, Bradford, United Kingdom, t.colak@bradford.ac.uk

[2] University of Bradford, Bradford, United Kingdom, r.s.r.qahwaji@brad.ac.uk

Abstract

The importance of real-time processing of solar data especially for space weather applications is increasing continuously. In this paper, we present an automated hybrid computer platform (ASAP) for the short-term prediction of significant solar flares using SOHO/MDI images. This system integrates image processing and machine learning to deliver these predictions. A machine learning-based system is designed to analyze years of sunspots and flares data to extract knowledge and to create associations that can be represented using computer-based learning rules. An imaging-based real time system that provides automated detection, grouping and then classification of recent sunspots based on the McIntosh classification is also created and integrated within this system. The properties of the sunspot regions are extracted automatically by the imaging system and processed using the machine learning rules to generate the real-time predictions. Several performance measurement criteria are used and the results are provided in this paper.

1 Introduction

Space weather is defined by the U.S. National Space Weather Program (NSWP) as "conditions on the Sun and in the solar wind, magnetosphere, ionosphere, and thermosphere

that can influence the performance and reliability of space-borne and ground-based technological systems and can endanger human life or health" (Koskinen et al., 2001). The importance of understanding of space weather is increasing because of the way solar activity affects life on Earth. We also rely more and more on communications and power systems, both vulnerable to space weather.

The most dramatic solar activity events affecting the terrestrial environment are solar flares and Coronal Mass Ejections (CMEs) (Pick et al., 2001). Solar flares and CMEs are solar eruptions that can spew vast quantities of radiation and charged particles into space (Lenz, 2004). The ability to predict major solar storms can give companies sufficient lead time to implement preventive measures (Lenz, 2004). Satellite operators, space agencies, aviation industry, power generation and distribution industry, oil and gas industry and railways can benefit from an effective space weather prediction system.

Solar activity is the driver of space weather. In order to predict solar activities we need to use real-time, high-quality data and data processing techniques (Wang et al., 2003). An automated space weather prediction service can be designed by combining solar physics with advanced image processing and machine learning techniques. Eruptions on the Sun, which are related to solar flares, travel to Earth in about 8 minutes in forms of light, radio waves, or X-rays. Proper warning of magnetic storms on Earth can be initiated if appropriate instruments to observe the Sun, the intervening space, and the Earth's magnetic field, are combined with efficient data processing techniques.

There are various research groups and organizations scattered around the world that are working on solar analysis and forecasting. These analyses and forecasts are subjective and depend mainly on the expert knowledge, but it is affected by fatigue and other human-related factors. On the other hand, objective computerized analysis of solar images can provide

automated processing and a very consistent performance by exploiting the large computational capabilities of modern computers to analyze and compare large amounts of current and historical data. However, the real challenge today is to design high-performance and accurate computer-based systems.

To our knowledge, there has been no fully automated system that can provide real-time prediction of significant solar flares that may affect our life on Earth. The accuracies of the previous semi-automated systems that have been designed are generally lower compared to the performance provided by subjective analysis. For example, WOLF (Miller, 1988) is an expert system that has been created to analyze active regions and sunspots and then predict the probability of solar flare occurrence. WOLF has a knowledge base consisting of a set of "if-then" rules and an inference engine which applies these rules. Manual user interaction is required to provide a description for the observed active region and sunspots. WOLF would then determine the McIntosh classification for the associated sunspot and hence, the probability of the described group producing a flare of specified X-ray intensity.

THEO (McIntosh, 1990) is another expert system that is also based on the McIntosh classification system, but includes information on spot growth, rotation and shear, and inferred magnetic topology. Both systems are not fully automated as they require user interaction and are not designed to handle solar images directly or automatically.

In this paper we introduce our fully automated pilot system called ASAP, which integrates image processing, machine learning and solar physics to provide automated prediction of solar flares based on the characteristics of sunspots. The prediction system is composed of two major stages: an image processing system and a machine learning system. The imaging system processes MDI continuum and magnetogram images in real-time mode to detect and classify sunspot groups and then determine their properties. On the other hand,

the machine learning system is trained using historical sunspots and flares data. Association algorithms are designed to associate flares with the sunspot groups that caused them and to create training sets that are used to train the learning algorithms and produce computerized learning rules. We will demonstrate the reliability of this system by testing it on recent solar images and comparing the generated predictions with the actual solar flares reported in solar catalogues.

This paper is organized as follows: Section 2 describes the imaging system that is responsible for the automated grouping and classification of sunspots, Section 3 describes the machine learning system that is trained on historical sunspots and flares data. The integration, performance and evaluation of the whole system are discussed in Section 4. The concluding remarks and suggestions for future work are presented in Section 5.

2 Sunspot Detection and Classification

A computer system that can automatically detect, group, and classify sunspots based on the McIntosh classification was presented in (Colak and Qahwaji, 2007). This system applies imaging techniques to SOHO/MDI continuum and magnetogram images to detect sunspot regions and extract their properties including their McIntosh classifications. In this work, this system for the first time is integrated with a machine learning-based system to provide real-time prediction for the possible occurrence of flares, which is described in the next section.

2.1 SOHO MDI images

The MDI instrument on SOHO provides almost continuous observations of the Sun in the white light continuum, in the vicinity of the Ni I 6767.8 Å photospheric absorption line. The instrument images the sun with a 1024x1024 CCD camera (Scherrer et al., 1995). White light pictures show how the Sun appears to the naked eye, while MDI continuum

(intensitygram) images are primarily used for sunspot observations. SOHO provides two to four MDI intensitygram images per day and twice as many magnetogram images with coverage which has been continuous since 1995.

MDI magnetogram images are used to measure the velocity and line-of-sight magnetic field strengths in the Sun's photosphere. The magnetogram images show the magnetic fields of the solar photosphere, with black and white areas indicating opposite magnetic polarities. The dark areas are regions of "south" magnetic polarity (pointing toward the Sun) and the white regions have "north" magnetic polarity (pointing outward). These images can be used for detecting active regions. In daily life magnetogram images are also used by observatories to study and cluster sunspot groups.

2.2 Summary of sunspot detection and classification algorithms

The detailed design and testing of this part of the system and description of algorithms are presented in (Colak and Qahwaji, 2007). Our automated sunspot detection, grouping and classification algorithm consists of three main stages: Pre-processing, sunspot detection and grouping, and sunspot classification. These stages can be summarized as follows:

Pre-processing of MDI images.

Stage-1 processing: Applied to both continuum and magnetogram images.

- Detect the solar disk, determine its radius and centre, create a mask and remove any information or marks (i.e., date and direction) from the image using the mask created.
- Calculate the Julian date and solar coordinates (The position angle, heliographic latitude, heliographic longitude) for the image using the equations in (Meeus, 1998).

Stage-2 processing: Applied only to magnetogram images.

- Map the magnetogram image from the Heliocentric-Cartesian coordinates to the Carrington Heliographic coordinates.
- Re-map the image to Heliocentric-Cartesian coordinates. Use centre, radius and solar coordinates of the continuum image as the new centre, radius and solar coordinates of the magnetogram image.

Sunspot grouping (Figure1).

- Detect sunspot candidates from MDI continuum images using intensity thresholding.
- Detect active region candidates from MDI magnetogram images using morphological image processing algorithms. The MDI magnetogram images show the footprint of the magnetic fields of the solar photosphere, with dark and light areas indicating opposite magnetic polarities. These dark and light areas are detected separately and combined afterwards to determine the active region candidates.
- Apply region growing to combine sunspot and active region candidates.
- Use neural networks to combine regions of opposite magnetic polarities in order to determine the exact boundaries of sunspot groups.
- Mark the detected sunspot groups.

McIntosh-based classification

- Extract local features from every sunspot in every group using image processing and neural networks.
    - Extract the length, height and area of the sunspot.

- Use neural networks to decide the type of penumbra (i.e., Mature or Rudimentary) and whether the sunspot is Symmetric or Asymmetric.

- Extract features from each sunspot group using image processing. The extracted features are length, largest spot, polarity and distribution.

- Apply all the extracted features to a decision tree to determine their McIntosh classification.

3 Solar Flare Prediction using Machine Learning

Solar flare research has shown that flares are mostly related to sunspots and active regions. A survey studying the reported associations between flares and sunspots in the solar physics literature is presented in (Qahwaji and Colak, 2007).

The system we introduce here extracts the expert knowledge embedded in historical data (i.e., solar catalogues) and represents it in computerized learning rules that enable computers to analyze recent solar data and provide reliable predictions. The implantation of this system is discussed below.

3.1 Knowledge Representation of Historical Data

Data from the publicly available sunspots group catalogue and the solar flares catalogue provided by the National Geophysical Data Centre (NGDC) are compared to find the associations between sunspots and flares. NGDC keeps records of data from several observatories around the world and holds one of the most comprehensive publicly available databases for solar features and activities. The NGDC flares catalogue provides information about dates, starting and ending times for flare eruptions, location, x-ray classification, and the National Oceanic and Atmospheric Administration (NOAA) number for the active regions

that are associated with the detected flares, while the NGDC sunspots catalogue provides information about their date, time, location, physical properties, and classification.

Both catalogues are investigated to associate significant flares with the sunspots that have caused them. All the reported flares and sunspots for the periods from 1st January 1982 till 31st December 2006, which includes 4595 (4258 M, and 337 X class) solar flares and 186324 sunspot groups are analyzed using the association algorithm described in (Qahwaji and Colak, 2007). The association algorithm is based on comparing the location (i.e. same NOAA number) and timing information (maximum 6 hours time difference between the erupting flare and its classified sunspot) of sunspot groups and solar flares. The association algorithm has managed to associate a total of 2346 (2156 M, and 190 X class) solar flares with their corresponding 2107 sunspot groups using their NOAA numbers and the six hours time difference. The difference between the number of solar flares and the number of associated sunspots is caused by the fact a sunspot group could produce more than one solar flare within the six hours time period. It is also worth mentioning that 4595 M and X class solar flares and 186324 sunspot groups are reported in the NGDC catalogues during this period. However, only 3575 (3272 M, and 305 X class) solar flares have location information. Also after eliminating the sunspot groups with missing or wrong entries in the sunspot catalogues the number of sunspot groups is reduced to 166719.

3.2 Flares Prediction system

The learning system introduced in this work is inspired by our previous work in (Qahwaji and Colak, 2007). Learning algorithms such as Neural Network (NN), Support Vector Machines (SVM) and Radial Bases Functions (RBF) were optimized, trained and then compared for flares predictions in (Qahwaji and Colak, 2007). In this work, features such as McIntosh classifications and daily sunspot numbers were used as inputs to the learning

algorithms. The training and generalization performances of the learning algorithms were evaluated using testing tools such as the Jack-knife technique (Fukunaga, 1990).

In (Qahwaji and Colak, 2007) the sunspot numbers were calculated using a solar cycle fitting model as suggested by (Hathaway et al., 1994). This was important then especially when dealing with historical data and trying to identify patterns of associations that could cause solar flare eruptions. However, this model is not quite suitable for real-time implementation as required by this work because the prediction of the solar cycle in the near future is not quite accurate. Therefore, in this paper we are using the area of sunspot groups together with the McIntosh classes as inputs to the learning algorithms in order to generate predictions for the M and X-class flares.

Our solar flare prediction is composed of two Neural Network (NN) systems that are working together as illustrated in Figure 2. The first NN uses the numerical representations of the three McIntosh classes for the sunspot under consideration together with its sunspot area as inputs. It generates the probability that this sunspot region is going to produce a significant solar flare in six hours time. Hence, our NN has four input nodes and one output node. This NN is trained using sunspot regions and solar flare associations as described above. The training vector for this NN contains numerical values representing the four inputs and their corresponding target. The target represents the actual "FLARE" / "NO FLARE" cases. This training vector is constructed as shown in Table 1. For example, if there is a sunspot region with a McIntosh classification of FKI and an area of 600 in millionths of solar hemisphere that is associated with a M or X-class flare then the training vector will be [0.9, 0.9, 0.5, 0.24; ***0.9***].

Table 1: Inputs and output values for first neural network that is used for determining the flaring probability.

Inputs				Output
McIntosh classes				
A= 0.10 H= 0.15 B= 0.30 C= 0.45 D= 0.60 E= 0.75 F= 0.90	X= 0 R=0.10 S=0.30 A=0.50 H=0.70 K=0.90	X=0 O=0.10 I=0.50 C=0.90	Normalized (with 2500) sunspot area	FLARE =0.9 NO FLARE= 0.1

When the first NN predicts that a flare is going to occur, the second NN is activated to determine whether the predicted flare is going to be M or X-class flare. The second NN is trained using a new training set that contains only the sunspot groups that were associated with X and M-class flares. Hence, this NN consists of four inputs and two outputs. The first output represents the M-class while the second represents the X-class flares. These outputs are assigned as described below:

If the sunspot group is associated with a M-class flare then the first output will be 0.9 and the second will be 0.1.

If the sunspot group is associated with a X-class flare then the first output will be 0.1 and the second will be 0.9.

If the sunspot group is associated with both M and X-class flares then the first and second outputs will be 0.9.

For example if there is a sunspot region with McIntosh classification of EKI and area of 500 in millionths of solar hemisphere that is associated only with M flare then the training vector will be [0.75, 0.9, 0.5, 0.20; ***0.9, 0.1***].

3.3 Optimization of the Neutral Networks Prediction System

The two neural networks are optimized by finding the minimum Mean Squared Error (MSE) during training for different NN topologies. MSE is calculated using Equation (1):

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(p_i - r_i)^2 \quad (1)$$

Where, n is the total number of examples in the training vector, p_i is the calculated value of each output for the inputs given in the training vector, and r_i is the real output value given in the training vector.

Several training experiments are carried out while changing the number of nodes in the hidden layer from 1 to 20. For every new experiment the MSE of the training is recorded and the number of hidden nodes with the least MSE is chosen. Both networks are optimized by using one hidden layer with ten nodes for the first NN and eleven nodes for the second NN.

4 Practical Implementation and Evaluation of the Hybrid System

The imaging and machine learning systems are integrated for the hybrid solar flares prediction system. The final system is shown in Figure 3. The complete integrated hybrid system provides automated prediction of solar flares from MDI images. The system starts its real-time operations by processing SOHO/MDI continuum and magnetogram images in the manner explained in Section 2 to provide automated McIntosh classifications for the detected sunspots. Then the McIntosh classified sunspots and their calculated areas are fed to the machine learning system described in Section 3 which is trained with 14 years of data after applying the association algorithm. Based on the embedded learning rules the system predicts if a major solar flare is going to occur or not. If a major solar flare is predicted then the probability of this solar flare to be M or X-class flare is also predicted. The entire system is

implemented in C++. It takes about 15 seconds to process the latest SOHO/MDI continuum and magnetogram images and generate these predictions.

4.1 Evaluation of the System

The performance of the hybrid system was evaluated by comparing the generated predictions with the actual flares activities as reported by NOAA SEC[1] in the NGDC X-ray solar flare catalogue. The system was tested on solar MDI intensitygram images from 1st March to 30th April 2001. This period corresponds to high solar activity that produced considerable number of M and X solar flares, which is important to test our system efficiently.

There were 241 MDI continuum (intensitygram) images available during this period that corresponds to four images per day. These MDI continuum images and their corresponding 241 MDI magnetogram images are processed using the hybrid system and a sunspot catalogue is created which we refer to as the Automated Sunspot Catalogue (ASC). Parts of ASC are shown in Table 2. All the automatically detected and classified sunspot groups and their flaring probabilities are provided. In addition to the ASC, the hybrid system represents its predictions visually as shown in Figure 4. The generated images show the detected sunspot groups and their flaring probabilities.

[1] http://www.sec.noaa.gov/

Table 2: The Automated Sunspot Catalogue (ASC). L (Length of the sunspot group), N (Number of sunspots within the group), F (significant Flaring), M(M-class flaring), X (X-class flaring), and PRO (Probability).

Date	Time	Location	CLS	L.	N.	Area	F.PRO.	M.PRO.	X.PRO.
01/04/2001	06:24	N25E25	HKX	1	4	134	0.1383	0	0
01/04/2001	06:24	N22W26	CSO	13	8	170	0.1126	0	0
01/04/2001	06:24	N16W45	FKC	28	25	1387	0.8147	0.8039	0.1206
01/04/2001	06:24	N11E22	CAO	4	7	56	0.0990	0	0
01/04/2001	06:24	N10W07	CAO	3	3	28	0.0975	0	0
01/04/2001	06:24	N09E29	AXX	1	0	8	0.0981	0	0
01/04/2001	06:24	S01W30	BXO	2	0	9	0.1047	0	0
01/04/2001	06:24	S05E04	DAO	17	6	128	0.1444	0	0
01/04/2001	06:24	S09W11	FSO	22	18	427	0.4202	0	0
01/04/2001	06:24	S07W29	DSO	6	7	89	0.1045	0	0
01/04/2001	06:24	S12W36	HRX	1	1	16	0.0927	0	0
01/04/2001	06:24	S12W75	BXO	2	6	68	0.1116	0	0
01/04/2001	06:24	S13E60	AXX	1	1	151	0.1457	0	0

In order to test the hybrid system we compared the solar flare predictions of each sunspot group in ASC to the x-ray solar flares reported in the NGDC solar flare catalogues. Between 1st March and 30th April 2001 there were 84 reported M and X class solar flares. Out of these 84 solar flares 75 of them were M-class and 9 of them were X- class flares.

In order to associate the sunspot groups detected by our algorithms with x-ray solar flares reported in the NGDC catalogue, we had to modify the association algorithm, introduced in (Qahwaji and Colak, 2007), to compare sunspots and solar flares based on their locations (latitude and longitude) instead of using their NOAA numbers. The new modifications are necessary because the sunspot groups in ASC do not have NOAA numbers. Also, not all the solar flares reported in NGDC flares catalogue have location information and the solar flares without location information are not included in this study. The association algorithm that we have used here is explained below:

- Read all the sunspot groups and their solar flare predictions as reported in ASC.

- Read the actual M and X-class flares with location information as reported in the NGDC solar flare catalogue.
- Carry out extensive search to associate each actual flare with its corresponding sunspot group by comparing their longitude, latitude, and time information. In order to confirm that a sunspot group and a solar flare are associated the following criteria must be met:
    - The difference between the longitude of the sunspot group and the solar flare must be less than 10 degrees.
    - The difference between the latitude of the sunspot group and solar flare must be less than 10 degrees.
    - The difference in time between the detected sunspot group and its associated flare should be less than six hours.
- If all the criteria are satisfied then highlight this sunspot group as associated with a solar flare, otherwise highlight it as not associated.

After the completion of the association process the prediction performance is evaluated using various verification measures as explained below.

4.2 Verification Results

The hybrid system generates predictions in numerical format, between 0.0 and 1.0, as shown in Table 2. A threshold value of 0.5 (50%) is used for determining the final predictions in this paper. If the output generated by the hybrid system is above this threshold value then a flare is predicted to occur. On the other hand, if the generated output is less than the threshold value then no flare is expected. In order to calculate the success of the generated predictions the association results are investigated using the following four criteria:

- If a sunspot group is associated with an actual flare, as explained previously, and a flares prediction is generated then this prediction is successful.
- If a sunspot group is associated with an actual flare but no flares prediction is generated then this prediction is not successful.
- If a sunspot group is not associated with any actual flare and no flares prediction is generated then this prediction is successful.
- If a sunspot group is not associated with any actual flare but a flares prediction is generated then this prediction is not successful.

For our testing period there exist 84 reported M and X-class flares in the NGDC catalogue. Only 63 (56 M- class and 7 X-class) have location information. These solar flares are compared with 2016 sunspot groups that were detected from 241 MDI image pairs and recorded in ASC.

Several forecast verification measures are used to evaluate our results as shown in table 3. The measures used are: Probability of Detection (POD), False Alarm Rate (FAR), Percent Correct (PC), Heidke Skill Score (HSS) and Quadratic Score (QR). These forecast measures are defined well in a recent paper by (Balch, 2008) and readers can refer to this paper for more information.

Table 3: Prediction evaluation measures

	Flare	**M Flare**	**X Flare**
Quadratic Score (QR)	0.055345	0.148737	0.158486
Probability of Detection (POD)	0.833333	0.814815	0.666667
False Alarm Rate (FAR)	0.769231	0.661538	0.969231
Percent Correct (PC)	0.939980	0.966270	0.966766
Heidke Skill Score (HSS)	0.363193	0.473472	0.090920

The quadratic score (QR) is simply the mean square error of the probabilities provided by the hybrid system. It is a measure to calculate the accuracy in probability predictions. A smaller QR indicates better accuracy in predictions. A perfect QR which is zero would require an accurate prediction of solar flare every time a solar flare occurs and an accurate prediction of no activity every time an event does not occur (Balch, 2008). The QR of our hybrid system for significant solar flare prediction, M-class predictions and X-class predictions are calculated 0.055345, 0.148737 and 0.158486 respectively. These values are encouraging especially for the significant solar flare prediction as it is close to zero.

POD measures the probability of actual solar flares being predicted correctly by the hybrid system. POD of a significant, M-class, and X-class solar flare predictions are 0.833333, 0.814815, and 0.666667, respectively. This means that the hybrid system predicted 83.3% of the significant solar flares correctly and this includes 81.4% of the M-class solar flares and 66.6% of the X-class flares correctly.

FAR measures the probability of the hybrid system predicting a solar flare that actually does not occur. FAR of a significant, M-class, and X-class solar flare are found to be 0.769231, 0.661538, and 0.969231, respectively. This means that 76.9% of the significant solar flares, 66.15% of the M-class flare, 96.15% of the X-class flare predictions by hybrid system were wrong.

PC is a measure showing the correct prediction rate of the overall system which is the ratio of successful flare and no flare predictions by our hybrid system. PC of significant, M-class, and X-class solar flare predictions are found to be 0.939980, 0.966270, and 0.966766 respectively. This means that the prediction rates of the hybrid system are very good when we take into account the correct flare and no flare predictions.

HSS is a measure showing the chance factor in predictions. The HSS can range from -1 (for no correct predictions) to +1 (for all correct predictions) and a value of zero is interpreted as the predictions been generated mainly by chance. HSS of a significant, M-class, and X-class solar flare are calculated 0.363193, 0.473472, and 0.090920 respectively. These numbers demonstrate that our system provides good performance for the significant flares and M-class predictions, but the performance of the X-class flares predictions is close to being produced by chance.

5 Conclusions

In this paper we have introduced, for the first time, an automated hybrid system called ASAP that integrates advanced machine learning and image processing techniques with solar physics to predict automatically whether a significant solar flare is going to erupt and the probability of this flare being X or M class flare.

The current results as presented in Section 4 are quite satisfactory but the performance of the hybrid system depends on the generalization capabilities of the machine learning system and the grouping and classification performance of the image processing system. For our hybrid system the HSS, POD, PC, and QR measures are quite good specially when predicting that a significant solar flare is going to erupt within six hours. However, the same thing cannot be said for FAR measure. This is caused mostly because of the six hours limit we used for predictions validity. There were some complex structured sunspot groups that can be tracked in several continues images and predicted to produce a significant flare by our hybrid system but this did not take place within the first six hours after detection, but rather in later stages. More than 90% of the sunspot groups that were predicted by the hybrid system to produce a significant flare actually produced a significant flare during their life time which is also good to determine if the group possesses a threat.

383 The prediction rates for the hybrid system can be improved by using the generalization
384 capability of machine learning system. Better generalization is obtained when more training
385 data are used. Hence, we believe that it is important to monitor the performance of the hybrid
386 system during its initial launch stages which includes comparing the prediction performance
387 with the actual flares reported by NOAA. Evolutionary algorithms may be used to allow the
388 learning algorithms to evolve and provide better optimization and generalization.

389 Therefore, ASAP is also designed to be web compliant and a real-time version is
390 available at http://spaceweather.inf.brad.ac.uk/index.html as a beta version since March 2007.
391 This system connects automatically to SOHO's website[2] to download the latest MDI
392 continuum and magnetogram images. After automatically running the algorithms described in
393 this paper the results are displayed on our website in a visual format as shown in Figure 4 and
394 always updated automatically.

395 The processing time for our system is approximately 15 seconds to process the two MDI
396 images and generate the predictions. This enables our system to provide near-real time
397 classifications and predictions. The system can be modified to accept other types of solar
398 images. Sometimes the characteristics of a sunspot group can change rapidly in hours and
399 cause flares, it is almost impossible for a human observer to determine what triggers this
400 process without having the appropriate tools. We believe that the system introduced here can
401 help researcher or observers to investigate and understand solar features and activities and to
402 identify patterns that could be useful for space weather predictions.

403 In our previous work (Qahwaji et al., 2007), (Qahwaji and Colak, 2007), we have shown
404 that SVM provides better generalization performance compared to NN. In the near future we
405 will integrate these findings and use SVM networks to enhance the generated predictions. We

[2] http://sohowww.nascom.nasa.gov/data/realtime/ Last Access: 2008

also believe designing a sunspots group tracking system that studies the evolution of the sunspot groups can improve the overall performance.

Acknowledgements

This work is supported by EPSRC (GR/T17588/01) and (**EP/F022948/1) grants**, which is entitled “Image Processing and Machine Learning Techniques for Short-Term Prediction of Solar Activity” and **"Image Processing, Machine Learning and Geometric Modeling for the 3D Representation of Solar Features", respectively.**

References

BALCH, C. C. (2008): Updated verification of the Space Weather Prediction Centre’s solar energetic particle prediction model. Space Weather 6,

COLAK, T. & QAHWAJI, R. (2007): Automated McIntosh-based classification of sunspot groups using MDI images. Topical Issues in Solar Physics,

FUKUNAGA, K. (1990): Introduction to Statistical Pattern Recognition. Academic Press, New York.

HATHAWAY, D., WILSON, R. M. & REICHMANN, E. J. (1994): The Shape of the Sunspot Cycle. Solar Physics 151, 177.

KOSKINEN, H., TANSKANEN, E., PIRJOLA, R., PULKKINEN, A., DYER, C., RODGERS, D. & CANNON, P. (2001): Space Weather Effects Catalogue. ESA Space Weather Program Feasibility Studies FMI, QinetiQ, RAL Consortium.

LENZ, D. (2004): Understanding and Predicting Space Weather. The Industrial Physicist 9/6, 18-21.

MCINTOSH, P. S. (1990): The Classification Of Sunspot Groups. Solar Physics 125/2, 251-267.

MEEUS, J. (1998): Astronomical Algorithms - Second Edition. Willmann-Bell, Inc., Virginia.

MILLER, R. W. (1988): Wolf - A Computer Expert System for Sunspot Classification And Solar-Flare Prediction. Journal of The Royal Astronomical Society Of Canada 82/4, 191-203.

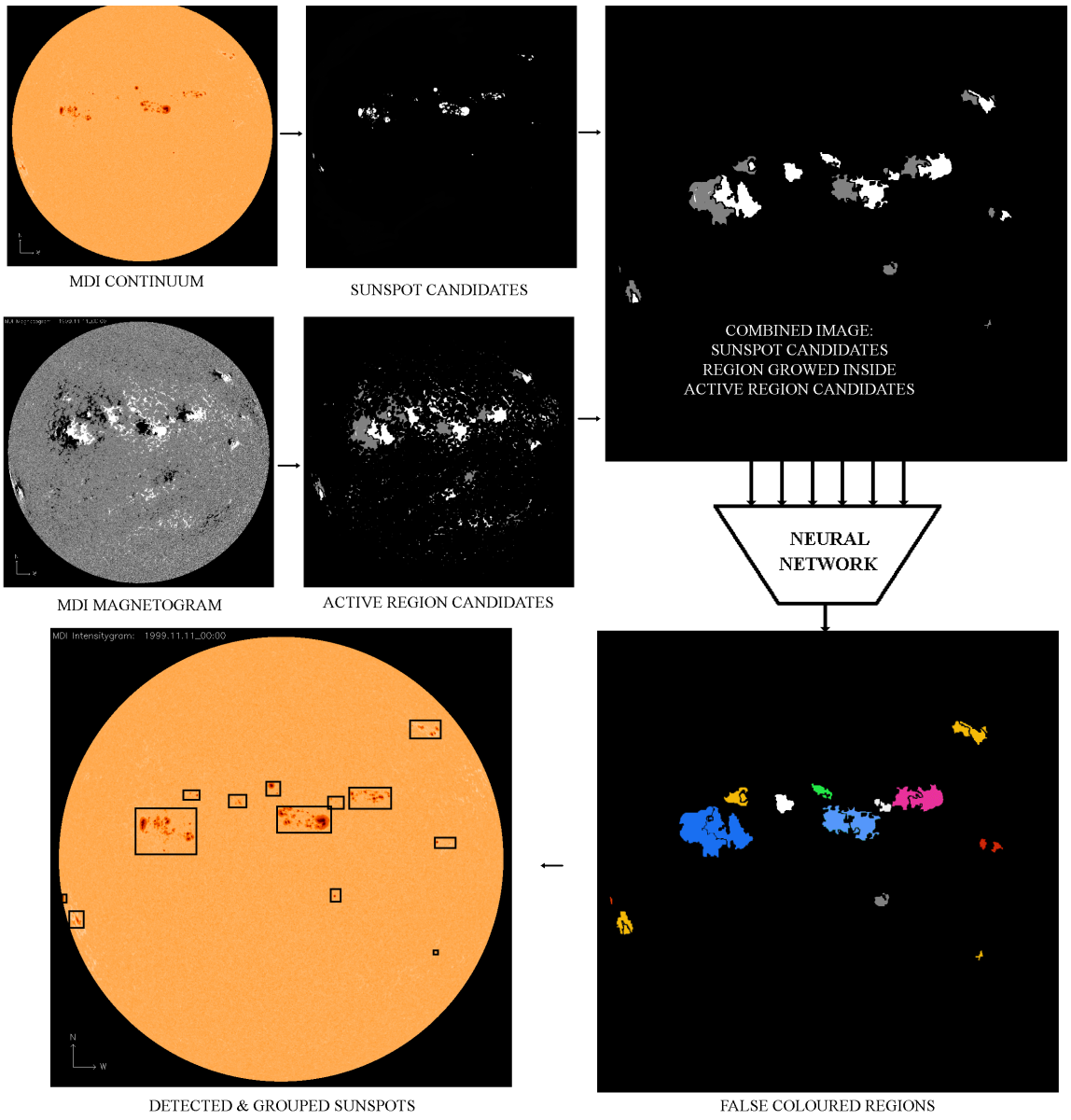
PICK, M., LATHUILLERE, C. & LILENSTEN, J. (2001): Ground Based Measurements. In: (ed.)^(eds): ESA Space Weather Program Feasibility Studies Alcatel-LPCE Consortium.

QAHWAJI, R. & COLAK, T. (2007): Automatic Short-Term Solar Flare Prediction Using Machine Learning and Sunspot Associations. Solar Physics 241/1, 195-211.

QAHWAJI, R., COLAK, T., AL-OMARI, M. & IPSON, S. (2007): Automated Machine Learning Based Prediction of CMEs Based on Flare Associations Topical Issues in Solar Physics,

SCHERRER, P. H., BOGART, R. S., BUSH, R. I., HOEKSEMA, J. T., KOSOVICHEV, A. G., SCHOU, J., ROSENBERG, W., SPRINGER, L., TARBELL, T. D., TITLE, A., WOLFSON, C. J., ZAYER, I., AKIN, D., CARVALHO, B., CHEVALIER, R., DUNCAN, D., EDWARDS, C., KATZ, N., LEVAY, M., LINDGREN, R., MATHUR, D., MORRISON, S., POPE, T., REHSE, R. & TORGERSON, D. (1995): The solar oscillations investigation - Michelson Doppler Imager. Solar Physics 162/1-2, 129-188.

WANG, H., QU, M., SHIH, F., DENKER, C., GERBESSIOTIS, A., LOFDAHL, M., REES, D. & KELLER, C. (2003): Innovative Information Technology for Space Weather Research. In: (ed.)^(eds): AAS 204th Meeting, Session 52 Virtual Observations in Solar Physics. The American Astronomical Society.



447

448 Figure 1: Stages and result of sunspot detection and grouping process.

449

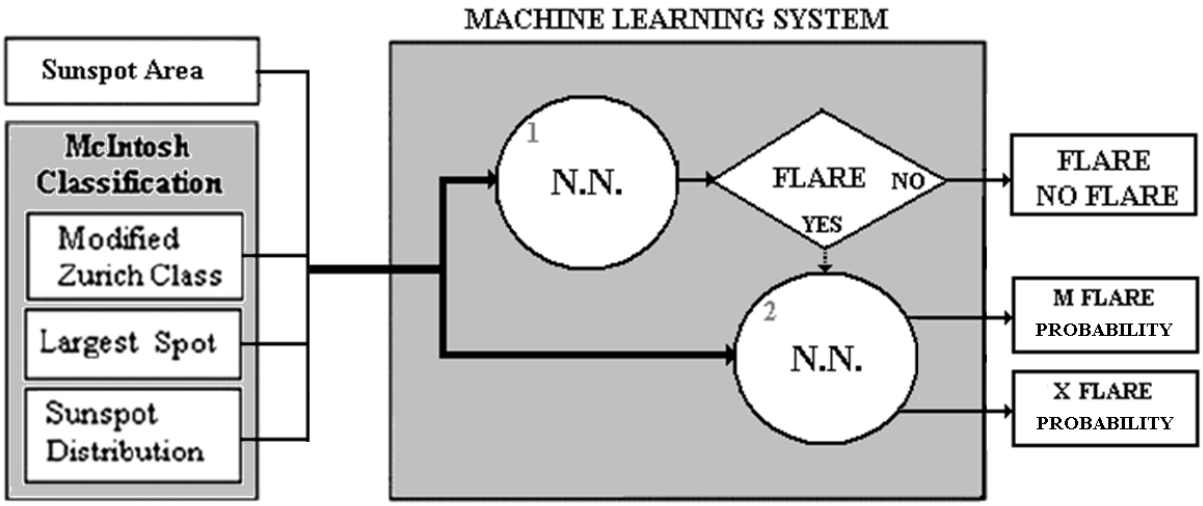


Figure 2: Machine learning system for flare prediction.

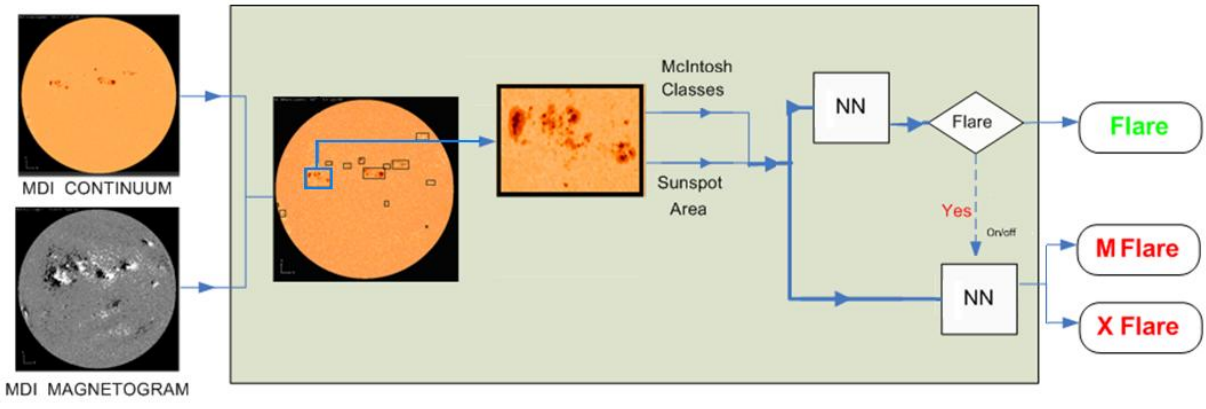
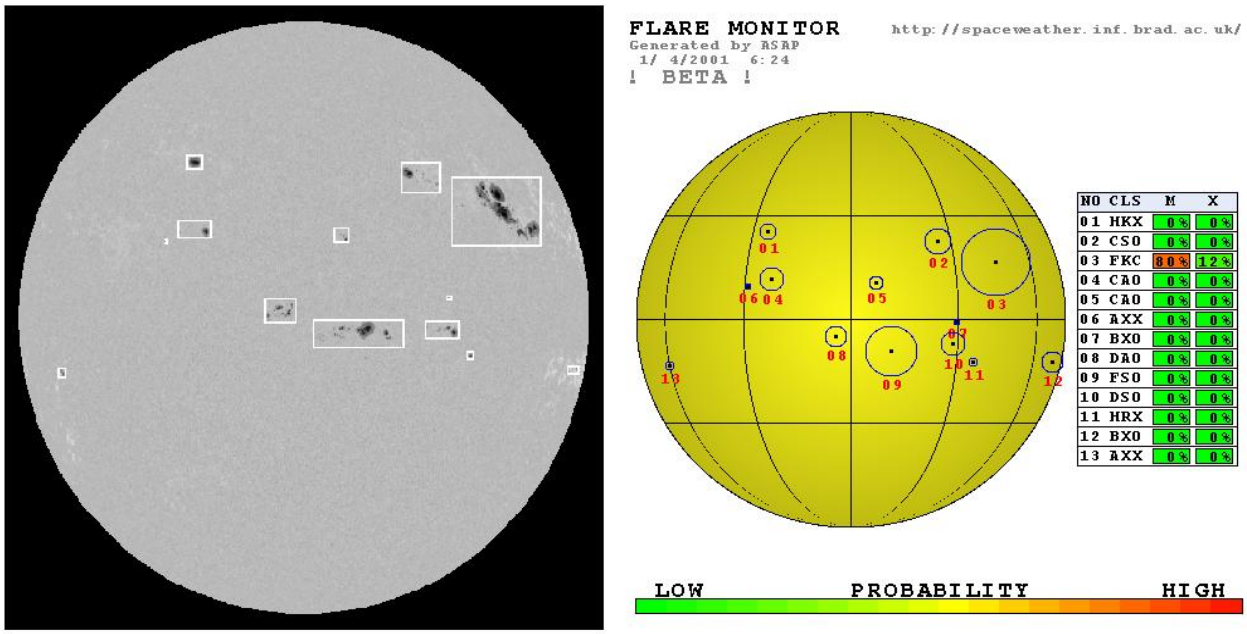


Figure 3: The final hybrid system



458

459 Figure 4: Two images generated by the hybrid system. The image to the left shows the
460 detected sunspot groups and the image to the right shows the classification and flaring
461 probabilities of the detected sunspot groups. The image to the right is updated automatically
462 on http://spaceweather.inf.brad.ac.uk/index.html every time a new MDI continuum image is
463 available on SOHO's website.