

# bradscholars

## Automated Solar Activity Prediction: A hybrid computer platform using machine learning and solar imaging for automated prediction of solar flares

Item Type	Article
Authors	Colak, Tufan;Qahwaji, Rami
Citation	Colak T, and Qahwaji R (2009) Automated Solar Activity Prediction: A hybrid computer platform using machine learning and solar imaging for automated prediction of solar flares. Space Weather. 7(6).
DOI	<a href="https://doi.org/10.1029/2008SW000401">https://doi.org/10.1029/2008SW000401</a>
Rights	© 2009 American Geophysical Union. Reproduced in accordance with the publisher's self-archiving policy.
Download date	2026-04-22 09:11:57
Link to Item	<a href="https://bradscholars.brad.ac.uk/handle/10454/4106.2">https://bradscholars.brad.ac.uk/handle/10454/4106.2</a>

## **The University of Bradford Institutional Repository**

<http://bradscholars.brad.ac.uk>

This work is made available online in accordance with publisher policies. Please refer to the repository record for this item and our Policy Document available from the repository home page for further information.

To see the final version of this work please visit the publisher's website. Where available access to the published online version may require a subscription.

Author(s): Colak, Tufan and Qahwaji, Rami S. R.

Title: Automated Solar Activity Prediction: A hybrid computer platform using machine learning and solar imaging for automated prediction of solar flares.

Publication year: 2009.

Journal title: Space Weather.

Publisher: American Geophysical Union.

Link to original published version:

<http://www.agu.org/pubs/crossref/2009/2008SW000401.shtml>

Citation: Colak, T., and Qahwaji, R. S. R. (2009). Automated Solar Activity Prediction: A hybrid computer platform using machine learning and solar imaging for automated prediction of solar flares. Space Weather, Vol. 7, No. 6.

Copyright statement: © 2009 American Geophysical Union. Reproduced in accordance with the publisher's self-archiving policy.

---

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

T. Colak<sup>1</sup> and R. Qahwaji<sup>2</sup>

[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

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

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

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

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

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

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

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

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

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

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

## 82 **2 Sunspot Detection and Classification**

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

### 89 **2.1 SOHO MDI images**

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

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

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

## 104 **2.2 Summary of sunspot detection and classification algorithms**

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

### 109 Pre-processing of MDI images.

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

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

115 Stage-2 processing: Applied only to magnetogram images.

116 ○ Map the magnetogram image from the Heliocentric-Cartesian coordinates to the  
117 Carrington Heliographic coordinates.

118 ○ Re-map the image to Heliocentric-Cartesian coordinates. Use centre, radius and solar  
119 coordinates of the continuum image as the new centre, radius and solar coordinates of the  
120 magnetogram image.

121 *Sunspot grouping (Figure1).*

122 ○ Detect sunspot candidates from MDI continuum images using intensity thresholding.

123 ○ Detect active region candidates from MDI magnetogram images using morphological  
124 image processing algorithms. The MDI magnetogram images show the footprint of the  
125 magnetic fields of the solar photosphere, with dark and light areas indicating opposite  
126 magnetic polarities. These dark and light areas are detected separately and combined  
127 afterwards to determine the active region candidates.

128 ○ Apply region growing to combine sunspot and active region candidates.

129 ○ Use neural networks to combine regions of opposite magnetic polarities in order to  
130 determine the exact boundaries of sunspot groups.

131 ○ Mark the detected sunspot groups.

132

133 *McIntosh-based classification*

134 ○ Extract local features from every sunspot in every group using image processing and  
135 neural networks.

136 ○ Extract the length, height and area of the sunspot.

- 137           ○ Use neural networks to decide the type of penumbra (i.e., Mature or Rudimentary)  
138           and whether the sunspot is Symmetric or Asymmetric.
- 139   ○ Extract features from each sunspot group using image processing. The extracted features  
140   are length, largest spot, polarity and distribution.
- 141   ○ Apply all the extracted features to a decision tree to determine their McIntosh  
142   classification.

### 143   **3   Solar Flare Prediction using Machine Learning**

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

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

#### 151   **3.1   Knowledge Representation of Historical Data**

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

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

161 Both catalogues are investigated to associate significant flares with the sunspots that  
162 have caused them. All the reported flares and sunspots for the periods from 1<sup>st</sup> January 1982  
163 till 31<sup>st</sup> December 2006, which includes 4595 (4258 M, and 337 X class) solar flares and  
164 186324 sunspot groups are analyzed using the association algorithm described in (Qahwaji  
165 and Colak, 2007). The association algorithm is based on comparing the location (i.e. same  
166 NOAA number) and timing information (maximum 6 hours time difference between the  
167 erupting flare and its classified sunspot) of sunspot groups and solar flares. The association  
168 algorithm has managed to associate a total of 2346 (2156 M, and 190 X class) solar flares  
169 with their corresponding 2107 sunspot groups using their NOAA numbers and the six hours  
170 time difference. The difference between the number of solar flares and the number of  
171 associated sunspots is caused by the fact a sunspot group could produce more than one solar  
172 flare within the six hours time period. It is also worth mentioning that 4595 M and X class  
173 solar flares and 186324 sunspot groups are reported in the NGDC catalogues during this  
174 period. However, only 3575 (3272 M, and 305 X class) solar flares have location information.  
175 Also after eliminating the sunspot groups with missing or wrong entries in the sunspot  
176 catalogues the number of sunspot groups is reduced to 166719.

### 177 **3.2 Flares Prediction system**

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

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

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

193 Our solar flare prediction is composed of two Neural Network (NN) systems that are  
194 working together as illustrated in Figure 2. The first NN uses the numerical representations of  
195 the three McIntosh classes for the sunspot under consideration together with its sunspot area  
196 as inputs. It generates the probability that this sunspot region is going to produce a significant  
197 solar flare in six hours time. Hence, our NN has four input nodes and one output node. This  
198 NN is trained using sunspot regions and solar flare associations as described above. The  
199 training vector for this NN contains numerical values representing the four inputs and their  
200 corresponding target. The target represents the actual “FLARE” / “NO FLARE” cases. This  
201 training vector is constructed as shown in Table 1. For example, if there is a sunspot region  
202 with a McIntosh classification of FKI and an area of 600 in millionths of solar hemisphere  
203 that is associated with a M or X-class flare then the training vector will be [0.9, 0.9, 0.5, 0.24;  
204 **0.9**].

205

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

Inputs			Output
McIntosh classes			Normalized (with 2500) sunspot area
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	
			FLARE =0.9 NO FLARE= 0.1

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

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

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

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

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

### 223 3.3 Optimization of the Neutral Networks Prediction System

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

$$226 \text{MSE} = \frac{1}{n} \sum_{i=1}^n (p_i - r_i)^2 \quad (1)$$

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

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

### 234 4 Practical Implementation and Evaluation of the Hybrid System

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

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

#### 247 **4.1 Evaluation of the System**

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

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

262

263

264

---

<sup>1</sup> <http://www.sec.noaa.gov/>

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

268	Date	Time	Location	CLS	L.	N.	Area	F.PRO.	M.PRO.	X.PRO.
269	01/04/2001	06:24	N25E25	HKX	1	4	134	0.1383	0	0
270	01/04/2001	06:24	N22W26	CSO	13	8	170	0.1126	0	0
271	01/04/2001	06:24	N16W45	FKC	28	25	1387	0.8147	0.8039	0.1206
272	01/04/2001	06:24	N11E22	CAO	4	7	56	0.0990	0	0
273	01/04/2001	06:24	N10W07	CAO	3	3	28	0.0975	0	0
274	01/04/2001	06:24	N09E29	AXX	1	0	8	0.0981	0	0
275	01/04/2001	06:24	S01W30	BXO	2	0	9	0.1047	0	0
276	01/04/2001	06:24	S05E04	DAO	17	6	128	0.1444	0	0
277	01/04/2001	06:24	S09W11	FSO	22	18	427	0.4202	0	0
278	01/04/2001	06:24	S07W29	DSO	6	7	89	0.1045	0	0
279	01/04/2001	06:24	S12W36	HRX	1	1	16	0.0927	0	0
280	01/04/2001	06:24	S12W75	BXO	2	6	68	0.1116	0	0
281	01/04/2001	06:24	S13E60	AXX	1	1	151	0.1457	0	0

282

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

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

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

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

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

## 311 **4.2 Verification Results**

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

- 318       • If a sunspot group is associated with an actual flare, as explained previously, and a  
319       flares prediction is generated then this prediction is successful.
- 320       • If a sunspot group is associated with an actual flare but no flares prediction is  
321       generated then this prediction is not successful.
- 322       • If a sunspot group is not associated with any actual flare and no flares prediction is  
323       generated then this prediction is successful.
- 324       • If a sunspot group is not associated with any actual flare but a flares prediction is  
325       generated then this prediction is not successful.

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

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

335       Table 3: Prediction evaluation measures

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

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

344           POD measures the probability of actual solar flares being predicted correctly by the  
345 hybrid system. POD of a significant, M-class, and X-class solar flare predictions are  
346 0.833333, 0.814815, and 0.666667, respectively. This means that the hybrid system predicted  
347 83.3% of the significant solar flares correctly and this includes 81.4% of the M-class solar  
348 flares and 66.6% of the X-class flares correctly.

349           FAR measures the probability of the hybrid system predicting a solar flare that  
350 actually does not occur. FAR of a significant, M-class, and X-class solar flare are found to be  
351 0.769231, 0.661538, and 0.969231, respectively. This means that 76.9% of the significant  
352 solar flares, 66.15% of the M-class flare, 96.15% of the X-class flare predictions by hybrid  
353 system were wrong.

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

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

## 366 **5 Conclusions**

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

371 The current results as presented in Section 4 are quite satisfactory but the performance of  
372 the hybrid system depends on the generalization capabilities of the machine learning system  
373 and the grouping and classification performance of the image processing system. For our  
374 hybrid system the HSS, POD, PC, and QR measures are quite good specially when predicting  
375 that a significant solar flare is going to erupt within six hours. However, the same thing  
376 cannot be said for FAR measure. This is caused mostly because of the six hours limit we used  
377 for predictions validity. There were some complex structured sunspot groups that can be  
378 tracked in several continues images and predicted to produce a significant flare by our hybrid  
379 system but this did not take place within the first six hours after detection, but rather in later  
380 stages. More than 90% of the sunspot groups that were predicted by the hybrid system to  
381 produce a significant flare actually produced a significant flare during their life time which is  
382 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<sup>2</sup> 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

---

<sup>2</sup> <http://sohowww.nascom.nasa.gov/data/realtime/> Last Access: 2008

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

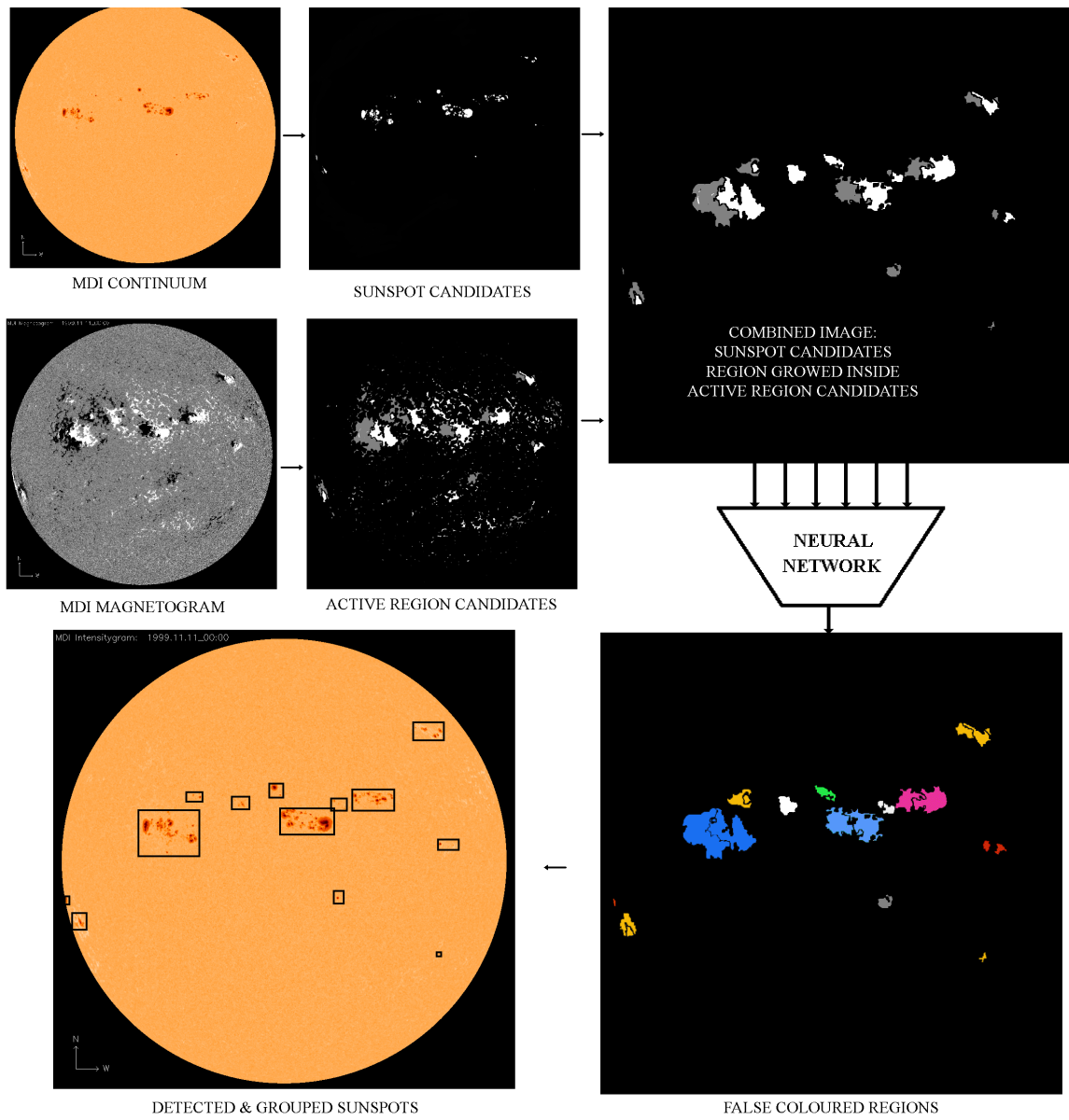
408

## 409 **Acknowledgements**

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

## 414 **References**

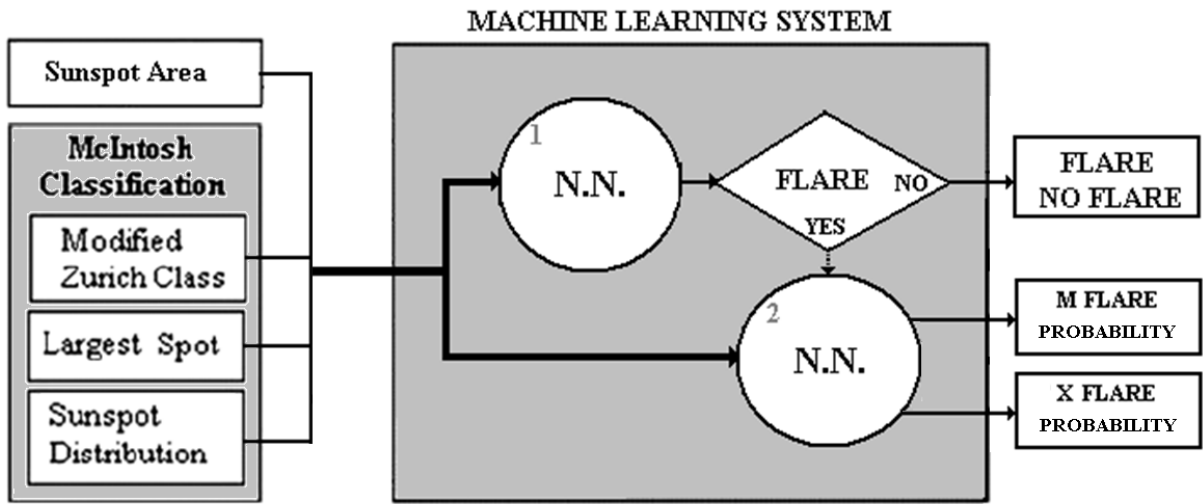
- 415 BALCH, C. C. (2008): Updated verification of the Space Weather Prediction Centre’s solar energetic particle  
416 prediction model. *Space Weather* 6,  
417 COLAK, T. & QAHWAJI, R. (2007): Automated McIntosh-based classification of sunspot groups using MDI  
418 images. *Topical Issues in Solar Physics*,  
419 FUKUNAGA, K. (1990): Introduction to Statistical Pattern Recognition. Academic Press, New York.  
420 HATHAWAY, D., WILSON, R. M. & REICHMANN, E. J. (1994): The Shape of the Sunspot Cycle. *Solar*  
421 *Physics* 151, 177.  
422 KOSKINEN, H., TANSKANEN, E., PIRJOLA, R., PULKKINEN, A., DYER, C., RODGERS, D. & CANNON,  
423 P. (2001): Space Weather Effects Catalogue. ESA Space Weather Program Feasibility Studies FMI, QinetiQ,  
424 RAL Consortium.  
425 LENZ, D. (2004): Understanding and Predicting Space Weather. *The Industrial Physicist* 9/6, 18-21.  
426 MCINTOSH, P. S. (1990): The Classification Of Sunspot Groups. *Solar Physics* 125/2, 251-267.  
427 MEEUS, J. (1998): *Astronomical Algorithms - Second Edition*. Willmann-Bell, Inc., Virginia.  
428 MILLER, R. W. (1988): Wolf - A Computer Expert System for Sunspot Classification And Solar-Flare  
429 Prediction. *Journal of The Royal Astronomical Society Of Canada* 82/4, 191-203.  
430 PICK, M., LATHUILLERE, C. & LILENSTEN, J. (2001): Ground Based Measurements. In: (ed.)^(eds): ESA  
431 Space Weather Program Feasibility Studies Alcatel-LPCE Consortium.  
432 QAHWAJI, R. & COLAK, T. (2007): Automatic Short-Term Solar Flare Prediction Using Machine Learning  
433 and Sunspot Associations. *Solar Physics* 241/1, 195-211.  
434 QAHWAJI, R., COLAK, T., AL-OMARI, M. & IPSON, S. (2007): Automated Machine Learning Based  
435 Prediction of CMEs Based on Flare Associations *Topical Issues in Solar Physics*,  
436 SCHERRER, P. H., BOGART, R. S., BUSH, R. I., HOEKSEMA, J. T., KOSOVICHEV, A. G., SCHOU, J.,  
437 ROSENBERG, W., SPRINGER, L., TARBELL, T. D., TITLE, A., WOLFSON, C. J., ZAYER, I., AKIN,  
438 D., CARVALHO, B., CHEVALIER, R., DUNCAN, D., EDWARDS, C., KATZ, N., LEVAY, M.,  
439 LINDGREN, R., MATHUR, D., MORRISON, S., POPE, T., REHSE, R. & TORGERSON, D. (1995): The  
440 solar oscillations investigation - Michelson Doppler Imager. *Solar Physics* 162/1-2, 129-188.  
441 WANG, H., QU, M., SHIH, F., DENKER, C., GERBESSIOTIS, A., LOFDAHL, M., REES, D. & KELLER, C.  
442 (2003): Innovative Information Technology for Space Weather Research. In: (ed.)^(eds): AAS 204th  
443 Meeting, Session 52 Virtual Observations in Solar Physics. The American Astronomical Society.  
444  
445  
446



447

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

449



450

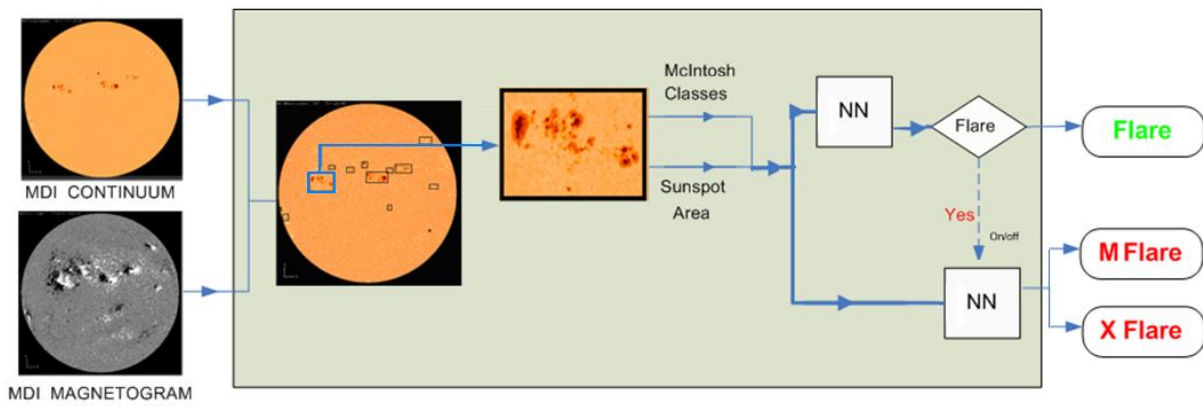
451 Figure 2: Machine learning system for flare prediction.

452

453

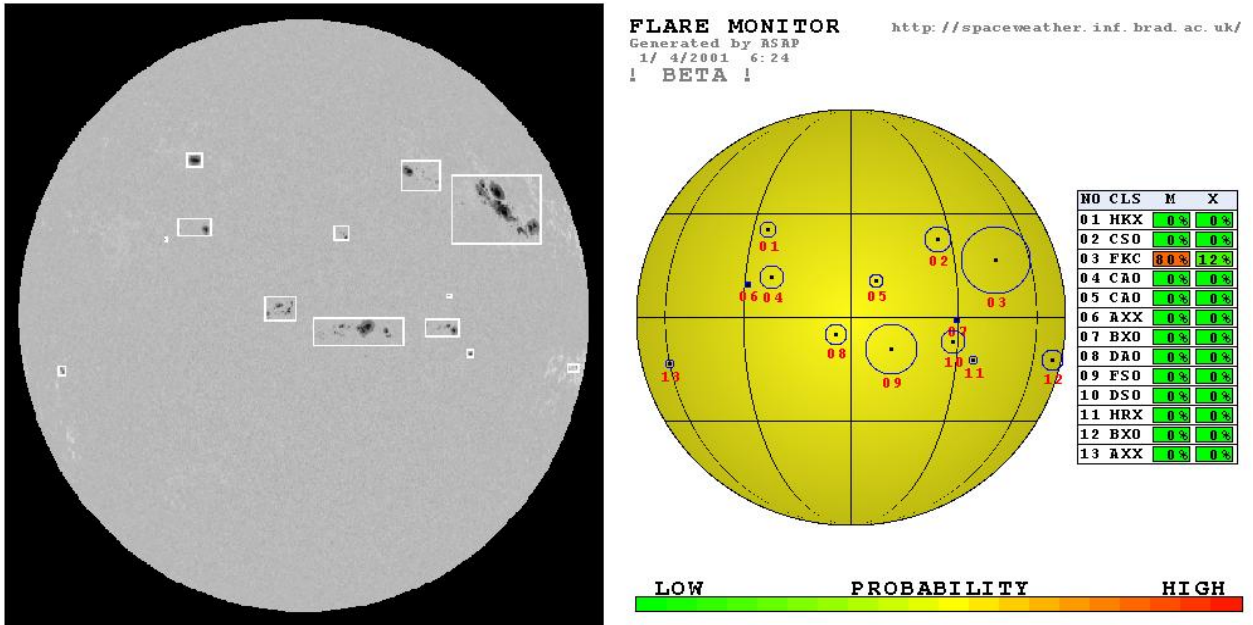
454

455



456

457 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.