AUTOMATIC SOLAR FLARE DETECTION USING NEURAL NETWORK TECHNIQUES

ROBERTO A. FERNANDEZ BORDA[1], PABLO D. MININNI[2], CRISTINA H. MANDRINI[1], DANIEL O. GÓMEZ[1,*], OTTO H. BAUER[3] and MARTA G. ROVIRA[1]

[1]*Instituto de Astronomía y Física del Espacio, C.C.67, Suc. 28, 1428 Buenos Aires, Argentina*
[2]*Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, 1428 Buenos Aires, Argentina*
[3]*Max Planck Institute für Extraterrestrische Physik, Germany*

(Received 27 August 2001; accepted 31 October 2001)

Abstract. We present a new method for automatic detection of flare events from images in the optical range. The method uses neural networks for pattern recognition and is conceived to be applied to full-disk Hα images. Images are analyzed in real time, which allows for the design of automatic patrol processes able to detect and record flare events with the best time resolution available without human assistance. We use a neural network consisting of two layers, a hidden layer of nonlinear neurodes and an output layer of one linear neurode. The network was trained using a back-propagation algorithm and a set of full-disk solar images obtained by HASTA (Hα Solar Telescope for Argentina), which is located at the Estación de Altura Ulrico Cesco of OAFA (Observatorio Astronómico Félix Aguilar), El Leoncito, San Juan, Argentina. This method is appropriate for the detection of solar flares in the complete optical classification, being portable to any Hα instrument and providing unique criteria for flare detection independent of the observer.

1. Introduction

The development of an automatic procedure to detect solar flares is a challenging task, since it is very difficult to take into account the complexity involved in the human classification process. For instance, human detection is strongly dependent on the observer's capabilities. In principle, a human recognition process can be reduced to a well-defined procedure or recipe using linear algorithms. However, the machine should be told in advance, and in great detail, the exact series of steps required to perform the algorithm. Also, the corresponding set of data should be given in a very precise format.

On the other hand, neural networks are simple algorithms that can learn from a human training set. They are robust in the presence of noise, and can deal with previously unrecognized patterns and generalize from the training set. In this way, nets are good at perceptual tasks and associative recall. A neural network is a system

*Also at Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires.

Solar Physics **206:** 347–357, 2002.

designed to recognize patterns in a given set of inputs. Basically, it is an interconnected assembly of simple processing elements, units or nodes, whose functionality was originally inspired on the animal neuron. The architecture consists of an input layer, a number of hidden layers, and an output layer. The processing ability of the network is stored in the inter-unit connection weights, obtained by a process of learning from a set of training patterns.

These kinds of algorithms based on a given number of weights, which are optimized from a training set, have been used for a long time in astronomy. The sunspot number is a clear example (Foukal, 1990). In 1848 the Swiss astronomer Wolf, using data collected since the 17th century, introduced the criteria used until today to measure the sunspot number. He defined the sunspot index number as $R = k(10g + f)$, where f is the number of individual spots, g the number of recognizable spot groups, and k is a correction factor that is intended to adjust for differences between observers. He established an international network of observatories, each with its own weight k. Note also that the number of groups has a weight of 10 against the number of individual spots. The most remarkable feature seen in this index is that, in spite of its arbitrary definition, it is a very reliable indicator of the Sun's magnetic field activity. Also it is approximately proportional to the total area covered by sunspots and to the energy flux emitted by the Sun at the wavelength of 10.5 cm.

When used for pattern recognition, the goal is to produce an output which is true if the pattern is present in the analyzed image, and false in any other case. One of the challenges of using a neural network for pattern recognition is to determine the relevant input variables (Looney, 1996; Therrien, 1989). Analyzing full images takes a huge amount of memory and CPU time, so a number of features which are believed to contain sufficient information to recognize the pattern is extracted from the image using a linear algorithm. This features are used as the input vector for the neural network.

In this paper we present a new method to automatically detect flare events in the optical range. This method uses neural networks for pattern recognition and is conceived to be applied to full-disk raw Hα images. Images are pre-analyzed in real time, which allows for the design of automatic patrol processes, which can detect and record flare events with the best time resolution achievable by the instrument, without human assistance. Raw images are used because flat field and dark corrections would consume valuable computing time. On the other hand, this extra computing time is unnecessary most of the time, since one of the main advantages of neural networks is that they can learn to recognize features in noisy datasets.

In spite of its potential importance, only a few applications of neural network techniques have been developed to perform automatic detection of solar-related events (Gothoskar and Khobragade, 1995) to detect solar wind activity using power spectra at radio frequencies (Wang, 2000) to detect and classify proton events. As far as we know, this paper presents the first application of neural networks to detect solar events in optical images.

We have applied this method to full-disk raw images obtained by HASTA (Hα Solar Telescope for Argentina), which is located at OAFA (Observatorio Astronómico Félix Aguilar), El Leoncito, San Juan, Argentina. The method is appropriate for the detection of flares in the complete optical classification, being portable to any Hα instrument and providing unique criteria for flare detection independent of the observer.

Section 2 summarizes the technical characteristics of the HASTA telescope. In Section 3 we describe the architecture of the neural network used for the present analysis, and the back-propagation training algorithm, and in Section 4 we enumerate the steps followed to build the training set. In Section 5 we show the results and test our method using real-time observations made by HASTA. Finally, in Section 6 we summarize the main results obtained in the present paper.

2. The Instrument

The HASTA instrument, which had its first light in May 1998, was designed to observe the Sun in the Hα spectral line, with good spatial resolution and a very high cadence. Because of these characteristics, it is an excellent instrument to support the analysis and modeling of flares and flare-related phenomena.

HASTA is located at El Leoncito, in the province of San Juan, Argentina (longitude: $-69.3°$, latitude: $-31.8°$, altitude : 2370 m) and it consists of three technical units: the telescope, the CCD camera and the computer (see Figure 1).

The telescope is the so-called 'Triple Solar Telescope', which was operated by the University of München at Wendelstein Observatory in the Bavarian Alps until 1987. Although only one of its three tubes is being presently used, the other two could be used to obtain images in other spectral wavelengths. This tube is a refractor telescope with an aperture of 110 mm of diameter and a focal length of 165 cm. The primary lens collimates the sunlight, which goes through an attenuation filter and then enters the Hα Lyot filter. The Lyot is a tunable filter optimized to perform measurements at the center of the Hα line (6563 Å) with a bandwidth of 0.3 Å and tuning range of 1 Å.

The telescope was built by Carl Zeiss Co. in the 1950s and in the 1990s was adapted to make full-disk Hα observations including the tunable Lyot filter and the CCD camera. The telescope's mount is equatorial and it has an automatic tracking system which allows to follow the Sun during an observing day. The Hα light is reflected at $90°$ by a mirror, which directs the beam toward the CCD camera. A camera lens set adjusts the full-disk image to the dimensions of the semiconductor CCD array. The final image diameter is 6.33 mm and the telescope diffraction limit at $\lambda = 6563$ Å is 1.5 arc sec, according to the Rayleigh criterion. The image scale is approximately 2.07 arc sec per pixel. Further details of these instruments have been reported elsewhere (Fernandez Borda *et al.*, 2002; Bagalá *et al.*, 1999).

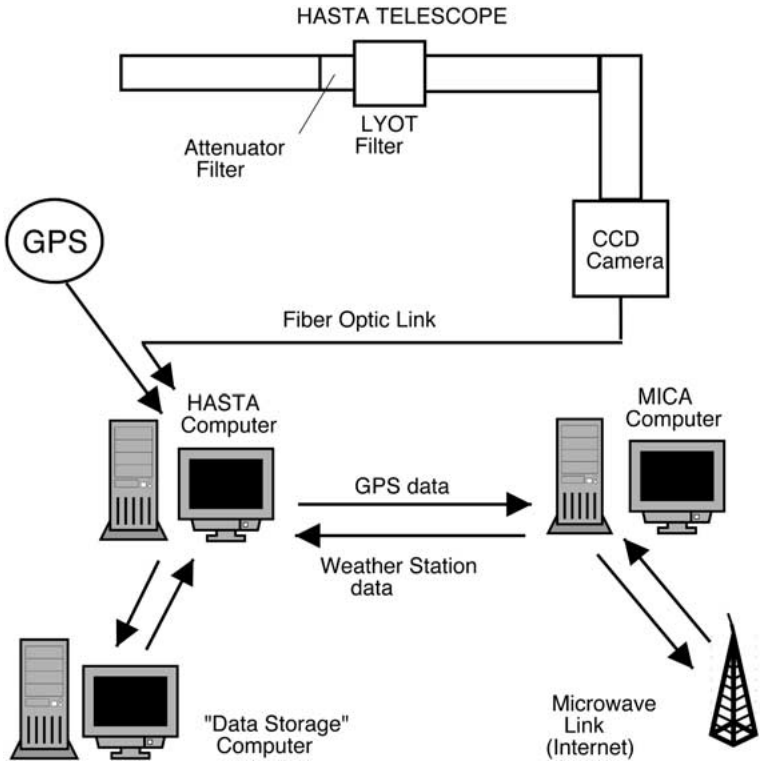

Figure 1. Logistic diagram of the HASTA instrument.

The CCD camera used for data acquisition is a SVGA-SensiCam with a CCD array of 1280 (horizontal) $\times$ 1024 (vertical) pixels. Each pixel is 6.7 μm $\times$ 6.7 μm and has a full-well capacity of 25000 e which allows to record all kind of impulsive events without saturation using intregration times within the range of 50–100 ms. The CCD presents a high quality level with a non-linearity of less than 1%.

The CCD camera is operated by a personal computer with a microprocessor Intel Pentium II of 233 MHz and 128 Mb of RAM memory. The camera is connected through an electronic interface to the PC and the data are downloaded to a set of SCSI hard disks. The images are finally stored in CDs. The Universal Time is provided through a GPS and the weather data are obtained from the weather station of the Mirror Coronagraph for Argentina (MICA, Stenborg *et al.*, 1999).

The instrument works in two different modes: the patrol mode, and the flare mode. In the patrol mode, HASTA takes full-disk images every 2 min. When a solar flare is detected, HASTA switches to the flare mode. In the flare mode, HASTA takes full-disk images every 10 s. This high cadence could be improved further, considering that the complete process of obtaining an image is approximately 2.2 s, which is the maximum image rate. To obtain a high-cadence sequence of Hα images right after the flare starts, a circulating buffer was developed. This buffer

holds twelve consecutive images with a cadence of 10 s, and is continuously being refreshed during the patrol mode. When the flare mode is activated, these twelve images are downloaded to the hard disk, which therefore provides two minutes of high-cadence images containing the initial stages of the flare.

3. Neural Networks and the Back-Propagation Algorithm

A neural node is a model of an artificial neuron. Broadly speaking, it has N synaptic inputs and a synaptic output. Each input line has an associated weight, which models the strength of the synapses. Signals in each input line are multiplied by its associated weight, and the weighted signals are added. The output of the neurode can be either 1 or -1 depending on whether or not the sum exceeds a given threshold value. Such a neurode, with linear response in its output, is known as a *perceptron*. Synaptic weights can be adjusted considering the effect of a given input to have the correct output, providing a schematic idea of a general learning rule (Looney, 1996, 1997; Werbos, 1994).

The output of the perceptron is a binary signal, since only one of two outcomes are possible. To extend the signal range to real numbers, activation functions are used in the output instead of a threshold. One function often used is the sigmoid

$$\sigma(s) = \frac{1 - e^{-as}}{1 + e^{-as}} = \tanh\left(\frac{as}{2}\right), \tag{1}$$

where σ is the output, s is the weighted sum over all the input nodes, and a controls the width of the sigmoid. Hereafter we set the value of a to $a = 1$, as is standard practice in the literature (Looney, 1997). We apply the term *nonlinear neurodes* to nodes with an activation function like this one to determine its output.

Our net is an example of a two layer network (see Figure 2). Inputs are applied to the left of the input layer, and the subsequent layers evaluate their node outputs until the output layer is reached. This process is referred to as a *forward pass*.

The input nodes are not artificial neurons, and act as branching nodes. As mentioned in the Introduction, this input layer is not fed directly with raw images. To speed up the classification and to make things manageable, the input nodes are fed with a set of parameters or *features*, obtained from pre-processing the images with a linear algorithm. This features are a set of N_I normalized real numbers which measure some properties of the images that are believed to contain sufficient information to recognize the pattern.

The hidden layer (the middle layer in Figure 2) contains N_H nonlinear neural nodes, with an activation function given by a sigmoid centered at zero. The output layer is a linear perceptron which is fed by the outputs of the hidden nodes and makes the final decision. Its threshold is fixed at zero and the output is -1 or 1 depending on whether the desired event has been observed or not.

To train this network, we use a supervised scheme. This means that a set of input feature vectors was constructed, together with the associated set of expected

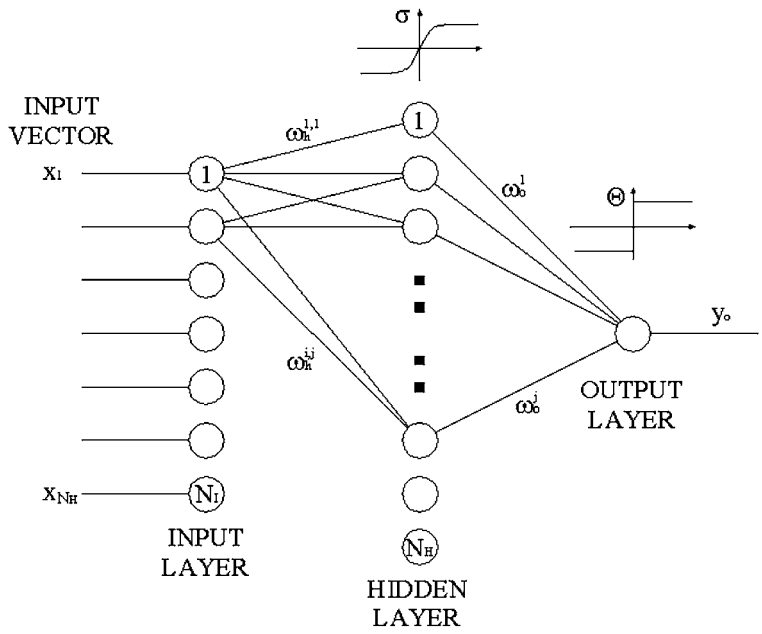

Figure 2. The architecture of our neural network.

output target vectors. An initial value for the weights is set. Then, the network input layer is recursively fed with the selected features, and the output is compared with the expected output targets. A difference in the result means that the net is misclassifying and the weights must be adjusted. The question is, can we train both layers (the hidden layer and the output node) in an iterative fashion? The idea is to perform a gradient descent on the obtained error, considered as a function of the weights in each node. To this end we use the *back-propagation training rule*.

In the *forward* pass, the neural network is fed with one of the input vectors and the final output y_o is calculated following

$$y_j = \sigma\left(\sum_{i=1}^{N_I} \omega_h^{i,j} x_i\right) \tag{2}$$

and

$$y_o = \Theta\left(\sum_{j=1}^{N_H} \omega_o^j y_j\right), \tag{3}$$

where y_j is the output at the jth hidden node, x_i is the input feature at the ith node, $\omega_h^{i,j}$ is the weight connecting the input node i with the hidden node j, and ω_o^j in the weight connecting the jth hidden node with the output node (see Figure 2). Θ is a step function with output $+1$ if its input is positive, and -1 in any other case.

If y_o differs from the expected output target t, the *backward* pass is started. First, the weights of the output node are adjusted using

$$\Delta\omega_o^j = \alpha_o \left(t - \sum_{j'=1}^{N_H} \omega_o^{j'} y_j \right) y_j \ , \tag{4}$$

where α_o is a free parameter which controls the learning rate of the output layer. Note that the change in the weight is proportional to the difference between the output of the node and the expected target, and the multiplication with y_j weights how much information enters by the jth input line.

Then, the weights of the hidden nodes are adjusted following

$$\Delta\omega_h{}^{i,j} = \alpha_h \sigma' \left(\sum_{i'=1}^{N_I} \omega_h^{i',j} x_i \right) \left(t - \sum_{j'=1}^{N_H} \omega_o^{j'} y_j \right) \omega_o^j x_i , \tag{5}$$

where α_h is the learning rate of the hidden layer and σ' is the derivative of the sigmoid function. Again, the change is proportional to the difference between the output of the node and the expected target. But in this case, σ' relates to how quickly the activation function can change the output (and therefore the error), and the multiplication with x_i weights the importance of the ith feature to determine the correct output.

This scheme is iterated for each feature presented at the input layer, until the error drops below a selected threshold.

4. The Training Set

To build the training set for the neural network, flares were selected by comparing the HASTA database from 1 January 1999 to 30 June 2000 against the Hα flare list in *Solar Geophysical Data* (*SGD*, Vol. 655–671, Part I). The selection criteria sample the complete optical flare classification (Švestka, 1976), all possible positions on the solar disk, different atmospheric conditions and common instrumental failures in order to avoid spurious detection of events.

Once both the topology of the network and the training algorithm have been decided, we need to define the set of features to conform the input vector. The selection of these features from the images is based on useful criteria used by observers as well as typical parameters that measure the observing conditions. Since the analysis of high-resolution and high-cadence images does obviously consume large amounts of memory and CPU time, it is essential to define a minimal set of features which contain all the relevant information. Our experience with the analysis of HASTA images, led us to decide the following set of seven features: (1) mean brightness of the frame, (2) standard deviation of the brightness, (3) pixel of maximum brightness derivative between consecutive images, (4) absolute brightness of that pixel, (5) radial position of that pixel, (6) variation of mean brightness between consecutive images (weather), and (7) contrast between the point with

maximum brightness derivative and its first neighbors. Features (1)–(4) are devoted to identify not just the flare, but more specifically, the image corresponding to the maximum activity for that particular flare. Feature (5), related to the distance of the flare to disk center, allows the network to take projection and limb darkening effects into account, without actually performing these corrections on the images, which would consume a non-negligible amount of CPU time. Cloudiness and other weather-related effects which can cause sudden variations in the total brightness of consecutive images are considered in feature (6). Finally, feature (7) is aimed at excluding dynamic spatial structures such as atmospheric turbulence or intense plage activity as possible flaring regions.

The vector of features is extracted from the raw images using a linear algorithm that we developed using IDL (Interactive Data Language, a package for data processing from Research Systems Inc.). This linear algorithm is efficient enough to allow the whole recognition process to run in real time. The identification of features plus the forward pass runs in approximately 5 s, which is faster than the integration time of the circulating buffer (approximately 10 s).

We also used the algorithm for feature identification to generate a GUI (Graphic User Interface) for the development of the training set (see Figure 3), which allows the training process of the neural network to become portable to other instruments. We made a careful choice of both positive and negative events, trying to cover a wide spectrum of patterns. Among the positive events, we included a wide variety of flares with different morphologies and intensities. Once the image corresponding to the maximum intensity for a given flare was identified (features (1)–(4)), we discarded all the remaining images associated with that flare from the dataset. As for the negative events, we included images with a total absence of flaring activity, images with plage activity, images with sunspots and other types of structures unrelated to flares.

From a total dataset of 361 events, we selected approximately 66% for the training set (237 events) and the remaining 33% (124 events) for the test set. One of the main advantages of neural network techniques is the relatively reduced datasets required to build appropriate selection criteria. An algorithmic or statistical approach would need a much larger number of examples in a very precise format, since noisy data would mislead the criteria. On the other hand, a neural network can deal with previously unseen patterns, and it is also robust in the presence of noise.

5. The Training Results

We set an error threshold in the training process of 3%, which was attained after approximately 10^3 iterations. Each iteration consists of a forward pass followed by a back-propagation step, which corrects the weights of all neural connections seek-

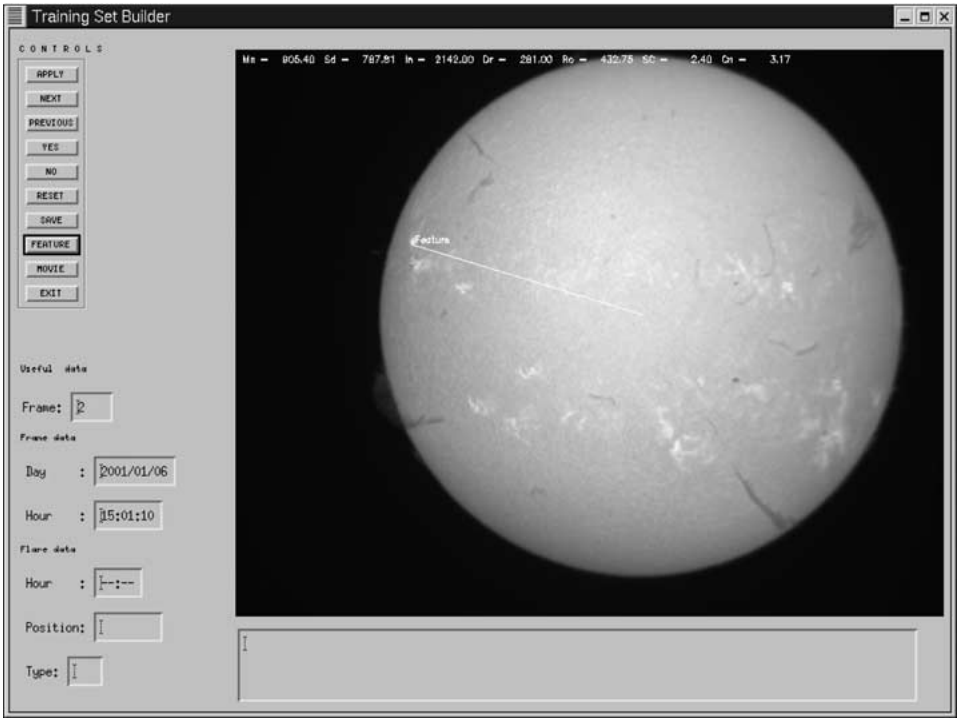

Figure 3. Graphic User Interface for the generation of the training set.

ing to minimize the error. The asymptotic values obtained for the weight factors, are the ones to be used to classify future events.

The order of events in the training set is shuffled randomly for each iteration, to reduce the risk of relaxing to local minima (Robbins and Monro, 1951). We confirm the strong dependence of the network's behavior with the number of nodes in the hidden layer. As suggested in the literature (Huang and Huang, 1991), we obtain that the most stable behavior and faster convergence arises for a number of nodes between N_i and $2N_i$, where $N_i = 7$ is the number of nodes at the input layer. We also find that the best results arise for learning rates (α) between 0.01 and 0.04. Therefore, we fix the number of nodes in the hidden layer as $N_h = 11$ and the learning rate as $\alpha = 0.02$.

After about 10^3 iterations for the training set of 237 events, we obtained the results depicted by Figure 4. The error is defined as the sum of the squares of the differences between the expected output target t and the actual output computed by the network, which has been properly normalized to represent the fraction of training errors. The reliability of the minimization process rests on the following aspects: (a) the monotonic decrease of the error, (b) the very low dispersion in the weight values once the error falls below the threshold, and (c) the independence of the final state from the initial weights. We tested this network with the test

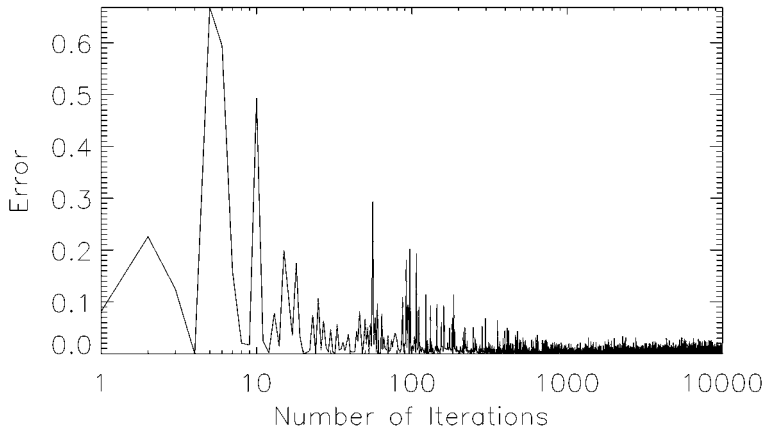

Figure 4. Error as a function of the number of iterations.

set of 124 events, and found that the fraction of misclassified events in this set was smaller than 5%. Considering a typical HASTA observing day of about eight hours, the mean number of events observed per day (considering its variation along the solar cycle), and the duration of individual events, such a low error implies a misclassified event every several days (two to five days).

6. Discussion

In the present paper we present a new method to automatically detect flare events in the optical range using neural networks for pattern recognition. It was designed to be applied to full-disk raw Hα images. We applied the method to images obtained by HASTA, using a training set of 237 events covering the period from 1 January 1999 to 30 June 2000.

After the training process was completed, the number of training errors was below 3%. The test set produced less than 5% of misclassified events. In a regular observing run, this means that an event will be misclassified every several days.

This method was proved useful for the automatic detection of flares in the complete range of Hα importance. The software associated with the neural network is friendly and easily portable to other observatories and instruments. Moreover, it is fast enough to run in real time, and provides a unique criterion for flare detection, which is independent of the observer.

Acknowledgements

We are grateful to an anonymous referee for his/her constructive comments that helped us to improve the original version of this manuscript. The authors also

want to express their gratitude to Drs Roberto Perazzo and Eduardo Honoré for very fruitful suggestions that guided the initial stages of this work. We acknowledge financial support from the University of Buenos Aires (grant TX065/98). The authors also acknowledge support from CONICET (grants PIP4536/96 and PIP4519/96). DOG, CHM and MGR are members of the Carrera del Investigador Científico (CONICET). RFB and PDM are fellows of CONICET. This study is based on data obtained at OAFA (El Leoncito, San Juan, Argentina) in the framework of the German–Argentinean HASTA/MICA Project, a collaboration of MPE, IAFE, OAFA and MPAe.

References

Bagalá, L. G., Bauer, O. H., Fernández Borda, R., Francile, C. Haerendel, G., Rieger, R., and Rovira, M. G.: 1999, *ESA* **SP-448**, 469.

Fernandez Borda, R., Francile, C., Bagalá, G., Bauer, O., Rovira, M. G., and Haerendel, G. 2002, *Solar Phys.*, in preparation.

Foukal, P.: 1990, *Solar Astrophysics*, Wiley Interscience Publ., New York.

Gothoskar, P. and Khobragade S.: 1995, *Monthly Notices Royal Astron. Soc.* **277**, 1274.

Huang, S. C. and Huang, Y. F.: 1991, *IEEE Trans. Neural Networks* **2**, 47.

Looney, C. G.: 1996, *IEEE Trans. Knowledge and Data Eng.* **8**, 211.

Looney, C. G.: 1997, *Pattern Recognition Using Neural Networks*, Oxford University Press, New York.

Robbins, H. and Monro, S.: 1951, *Ann. Math. Stat.* **22**, 400.

Stenborg, G., Schwenn, R., Srivastava, N., Inhester, B., Podlipnik, B., Rovira, M. G., and Francile, C.: 1999, *ESA* **SP-448**, 473.

Švestka, Z.: 1976, *Solar Flares*, D. Reidel Publ. Co., Dordrecht, Holland.

Therrien, W.: 1989, *Decision, Estimation and Classification*, Wiley, New York.

Wang, J. L.: 2000, *Chinese Astron. Astrophys.* **24**, 10.

Werbos, P. J.: 1994, *The Roots of Backpropagation*, Wiley, New York.