

A Comparative Evaluation of Automated Solar Filament Detection

M.A. Schuh · J.M. Banda · P.N. Bernasconi ·
R.A. Angryk · P.C.H. Martens

Received: 16 April 2013 / Accepted: 6 February 2014 / Published online: 27 February 2014
© Springer Science+Business Media Dordrecht 2014

Abstract We present a comparative evaluation for automated filament detection in H α solar images. By using metadata produced by the Advanced Automated Filament Detection and Characterization Code (AAFDCC) module, we adapted our trainable feature recognition (TFR) module to accurately detect regions in solar images containing filaments. We first analyze the AAFDCC module's metadata and then transform it into labeled datasets for machine-learning classification. Visualizations of data transformations and classification results are presented and accompanied by statistical findings. Our results confirm the reliable event reporting of the AAFDCC module and establishes our TFR module's ability to effectively detect solar filaments in H α solar images.

Keywords Automated feature finding · Filaments · Quantitative comparative evaluation

M.A. Schuh (✉) · J.M. Banda · R.A. Angryk
Department of Computer Science, Montana State University, 357 EPS, Bozeman, MT 59717-3880,
USA

e-mail: michael.schuh@cs.montana.edu

J.M. Banda

e-mail: juan.banda@cs.montana.edu

R.A. Angryk

e-mail: angryk@cs.montana.edu

P.N. Bernasconi

Applied Physics Laboratory, Johns Hopkins University, 11100 Johns Hopkins Road, Laurel, MD 20723,
USA

e-mail: pietro.bernasconi@jhuapl.edu

P.C.H. Martens

Department of Physics, Montana State University, 247 EPS, Bozeman, MT 59717-3880, USA

e-mail: martens@physics.montana.edu

P.C.H. Martens

Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA

1. Introduction

The stream of solar data produced by NASA's *Solar Dynamics Observatory* (SDO) mission is larger than all previous solar missions combined (Pesnell, Thompson, and Chamberlin, 2011). Traditional human-based analysis methods would be overwhelmed by the massive amount of potentially interesting SDO data (over 70000 images per day). Therefore, scientists are forced to rely on automated tools to efficiently process and analyze this non-stop data stream. The SDO Feature Finding Team (FFT; Martens *et al.*, 2012) is currently developing over a dozen of such automated modules to detect specific types of solar phenomena. Most of these are developed by specialized and independent teams responsible for only one type of pre-determined phenomenon.

As part of the SDO FFT, our interdisciplinary research group at Montana State University (Computer Science and Solar Physics) is developing a general purpose content-based image-retrieval (CBIR) system for solar imagery based on the metadata produced by our trainable feature recognition (TFR) module. Our CBIR system will allow users to explore the enormous archives of solar data through image-based searches rather than traditional text-based queries. This is a convenient and intuitive style of searching for images, because the user can simply provide an image of interest without requiring any necessary information (metadata) about the images being searched for. Not only does this make the solar image archives more accessible to all levels of expertise, but it also facilitates the rapid discovery of new phenomena that have unknown characteristics. No such method of discovery is currently available for solar physics image archives.

In this paper we investigate the use of our TFR module as a comparative analysis tool for other automated detection modules. We combine modules' metadata with a variety of machine-learning algorithms and evaluation techniques to create a supervised classification model to detect a specific type of event. Because the initial development of the TFR module occurred before the SDO launch, initial work was performed on image and event metadata from the *Transition Region And Coronal Explorer* (TRACE) mission (Handy *et al.*, 1999). As the incorporation of new phenomenon-specific modules continues (from the SDO FFT and elsewhere), we will have great opportunities to evaluate our effectiveness on new types of events reported by other automated modules. Each module we analyze allows us to measure our ability to distinguish and detect that module's specific event, and infer a potential baseline ability within our entire fully developed CBIR system in the future. While there is much more involved in creating an effective CBIR system, such as important invariant decisions surrounding query and retrieval, like incorporating event size and location into the definition of similarity, the fundamental ability to accurately detect an event of interest is an important first step.

The benefits of this ongoing work are three-fold:

- i) We gain detailed quantitative results about our (in-development) CBIR system's effectiveness at single-phenomenon detection (derived from the underlying TFR module).
- ii) The SDO Mission gains additional analysis and verification of the other FFT modules' results without investing in time-intensive human verification or the costly development of alternative (and essentially duplicate) modules.
- iii) The entire community gains more information and awareness about the SDO mission and the interdisciplinary achievements of the FFT through open discussion and dissemination of datasets, experiments, and results.

1.1. Contributions

We present an extension of our CBIR building framework (Banda, 2011) with the creation of additional components to facilitate comparative analyses using event metadata from automated detection modules. We report the first comparative evaluation results using the Advanced Automated Solar Filament Detection and Characterization Code (AAFDCC) module developed by Bernasconi, Rust, and Hakim (2005). While the AAFDCC module is officially part of the SDO FFT, it has been producing excellent results from ground-based observatories (H α images) for over ten years. These observatories capture only a few images each day, making human verification of the AAFDCC results a practical task – a luxury most other modules do not have. The availability of this large and already existing single-phenomenon dataset, and the likelihood of highly accurate metadata from numerous human verification, makes the AAFDCC module ideal for our preliminary investigations and allows us to establish a fairly confident benchmark for future module comparisons using SDO data.

The primary beneficiary of this work is the solar physics community. A standardized (and independent) method of comparative evaluation will verify reporting effectiveness for other automated modules. As human-based analysis of raw image data becomes increasingly prohibitive, the event repositories will become predominantly populated by automated findings, and these methods of data validation and comparative evaluation will become more important than ever before. Furthermore, any improvements to FFT modules, including our CBIR system, as a result of this work will directly enhance the quality of available tools for the solar physics community. Our evaluations will also shed further light on the specific image characteristics exclusive to solar imagery and solar events, contributing directly to the image processing and computer vision subfields of computer science. Lastly, this work will produce publicly available datasets (catalogs of event metadata) that others in the computer science or solar physics communities can readily use for future research.

The rest of the paper is outlined as follows. Section 2 contains background material relevant to the larger context of our work. Data preprocessing is described in Section 3, and we present experiments and results in Section 4. Finally, in Section 5 we review our conclusions and discuss directions for future work.

2. Background

Scientists have been acquiring and archiving satellite imagery of the Sun for over half a century, and the stream of images has been virtually uninterrupted for the past 20 years. Like in other domains, this data stream reflected the technology of the times, and it remained feasible for solar physicists to analyze images manually by identifying, labeling, and recording solar phenomena found in each image. The desire for increased spatial and temporal resolution – finer image detail and more frequent image capture – was first met with the 1996 launch of the ESA/NASA *Solar and Heliospheric Observatory* (SoHO; Domingo, Fleck, and Poland, 1995) providing more frequent full-disk images, and then with the 1998 launch of the TRACE providing more detailed partial-disk images (Handy *et al.*, 1999). These observations (along with others) vastly increased the data stream and revealed the emerging need to develop automated event detection methods which could replace traditional human analyses.

Given the current data stream of SDO (approximately 1.5 TB per day), brute-force human-based analysis is already quite impractical, and it will only get worse as future data streams will likely be even larger. This work highlights the growing need for interdisciplinary work between solar physicists and computer scientists to provide efficient and

automated algorithms for increasingly complex, large-scale data analysis, retrieval, and visualization. Furthermore, it emphasizes the importance and potential benefits of data mining and machine learning to the future of solar physics.

2.1. A CBIR System for Solar Images

A CBIR system facilitates image-based searching of image archives through analyzing the similarity of content within the images without the use of additional (text-based) metadata tied to each image. This allows a user to simply provide an image of interest as the search query and receive results of similar images in the archive, without any requirements of textual labels or domain expertise. CBIR systems have become very popular in domains with large image archives and insufficient labeling methods, such as medical image analysis, Earth sciences (geographic information systems; GIS), and online search contexts (*e.g.*, Google image search). Here we focus on the trainable feature recognition (TFR) module from our CBIR-related work, which uses supervised classification algorithms to identify different types of solar phenomena that it is trained on. Features within images are distinguished by the properties inherent to the image data, and a CBIR system is based primarily on the similarity of individual image representations (no metadata or event data). Often what then emerges is the ability to find similar looking phenomena (events, objects, *etc.*) from their intrinsically similar image parameter values. The first version of our solar CBIR system went live on 1 January 2013, and can be accessed at <http://cbsir.cs.montana.edu/sdocbir>. Several new versions are currently under research and development (Banda *et al.*, 2014; Schuh *et al.*, 2013a).

In previous work, Banda and Angryk (2010a, 2010b) evaluated a variety of possible numerical *image parameters* extracted from the raw image data, and the best ten were chosen based on their efficient processing time and classification accuracy (Banda and Angryk, 2010a, 2010b). It is well known that the effectiveness of image parameters greatly depends on the intrinsic characteristics of the images. Therefore, parameters were chosen that were empirically well suited for solar images. Preliminary phenomena classification evaluation was performed on human-labeled TRACE images to determine which image parameters best represented the phenomena. However, classification accuracy was balanced with the requirement of near real-time processing (parameter extraction from raw image data) as there is no chance to simply “catch up” to this continuous data stream. So while other parameters may have been more effective in classification results, they were still discarded if processing time was orders of magnitude larger than the others. Table 1 lists our ten chosen image parameters, and we direct the reader to Banda and Angryk (2010b) for more information on the selection process. We use the popular method of grid-based image segmentation which divides each image into smaller, equal-sized regions named cells. Image parameters are extracted from each individual cell, as shown in Figure 1.

2.2. Classification Algorithms

In general terms, classification algorithms attempt to find a hypothesis which best explains a given set of labeled data observations. In the context of our TFR module, a classification algorithm will build a classification model based on labeled training data in order to predict the most accurate labels for unseen testing data. We selected Naïve Bayes (NB), Support Vector Machines (SVM) with a linear kernel function, Decision Trees (C4.5), and Random Forests (RF) as our four possible classifiers. Linear classifiers, such as NB and SVM, create class separation through the linear combination of the data attributes. For example, suppose

Table 1 List of extracted image parameters, where L stands for the number of pixels in our image cell, z_i is the i -th pixel value, m is the mean, and $p(z_i)$ is the grayscale histogram representation of z at i . The fractal dimension is calculated based on the box-counting method where $N(\epsilon)$ is the number of boxes of side length ϵ required to cover the image cell. Labels are added for easier discussion.

Label	Name	Equation
P1	Entropy	$E = -\sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i)$
P2	Mean	$m = \frac{1}{L} \sum_{i=0}^{L-1} z_i$
P3	Standard deviation	$\sigma = \sqrt{\frac{1}{L} \sum_{i=0}^{L-1} (z_i - m)^2}$
P4	Fractal dimension	$D_0 = \lim_{\epsilon \rightarrow 0} \frac{\log N(\epsilon)}{\log \frac{1}{\epsilon}}$
P5	Third moment (skewness)	$\mu_3 = \sum_{i=0}^{L-1} (z_i - m)^3 p(z_i)$
P6	Fourth moment (kurtosis)	$\mu_4 = \sum_{i=0}^{L-1} (z_i - m)^4 p(z_i)$
P7	Uniformity	$U = \sum_{i=0}^{L-1} p^2(z_i)$
P8	Relative smoothness	$R = 1 - \frac{1}{1 + \sigma^2(z)}$
P9	Tamura contrast	*Algorithm omitted, see Tamura, Mori, and Yamawaki (1978)
P10	Tamura directionality	*Algorithm omitted, see Tamura, Mori, and Yamawaki (1978)

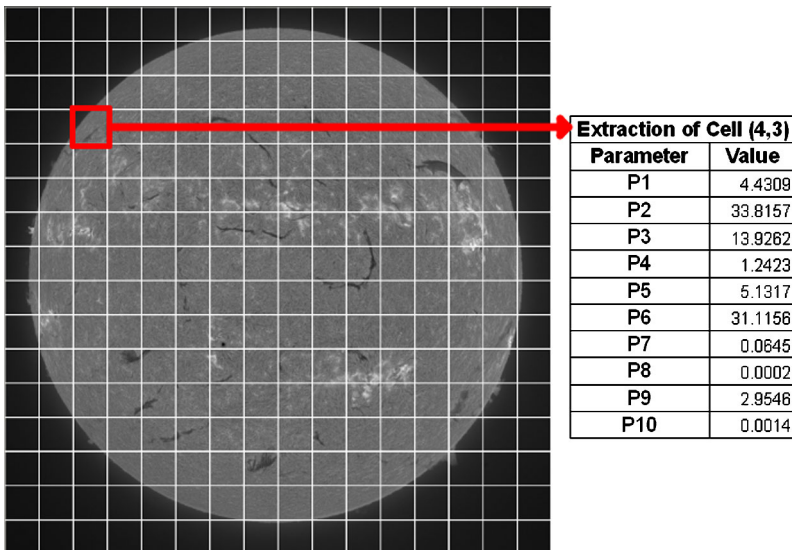


Figure 1 An example of our parameter extraction for a single cell in an image.

we have our ten parameters and two class labels (A and B), then a hypothesis could state: “If $(P1 \times P2) \geq x$ then class A, else class B.” In contrast, decision trees, such as C4.5 and RF, split the data into disjoint subsets on one attribute (dimension) at a time until adequate class separation has been achieved in the leaves of the tree. Here, an example hypothesis could state: “If $P1 \geq x$ then if $P2 \geq y$ then class A, else class B.” The tree root tests $P1$, which results in two leaves, one where $P1 \geq x$ and the other where $P1 < x$. Each of these leaves now recurse independently performing further tests and splits. In our example, the

Figure 2 Interpreting the results of a confusion matrix for binary classification.

		Classifier Prediction	
		Positive	Negative
AAFDCC Label	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

one leaf now checks if $P2 \geq y$, which branches into two new leaves, where all items are labeled class A in one and B in the other.

The Naïve Bayes classifier (Domingos and Pazzani, 1997) is surprisingly accurate in many applications and executes very fast, making it an ideal candidate for training on the massive number of images expected in the SDO repository. SVMs (Vapnik, 1995) have gained tremendous popularity in recent years due to their ability to maximize separation functions, which is said to improve the overall classifier accuracy on new data that it was not trained on. The main concern with applying SVMs to large-scale data is their slower training process, but we include them in our analysis for a more thorough comparison. The C4.5 classifier (Quinlan, 1986) is one of the most popular decision tree classifiers in the machine-learning community, and since it takes a greedy approach (*i.e.*, making the locally optimal choice at each step), it is also quick to compute. The RF classifier (Breiman, 2001), although slower to train, uses randomization to avoid local optima from poor greedy choices, providing a much more robust prediction than C4.5. It essentially creates many independent decision trees (hence a “forest”) and then uses a majority voting strategy to choose the best label based on the independent predictions of all trees in the forest.

A confusion matrix (or truth table) represents the possible outcomes of predicting a class label for a data instance. Since we have only two possible values for our class label (binary classification), there are four possible outcomes for each prediction by our classifiers. The four-element confusion matrix is shown in Figure 2, where each column of the matrix is the predicted class label by our classifier and each row is the actual class label, derived from the AAFDCC module. In our case the filament label is the *Positive* class and the non-filament label is the *Negative* class. Therefore, the four regions represent: true positive (TP), a filament cell labeled correctly as a filament; false negative (FN), a filament cell labeled incorrectly as a non-filament; false positive (FP), a non-filament cell labeled incorrectly as a filament; and true negative (TN), a non-filament cell labeled correctly as a non-filament. Among other things, we report the classification accuracy and precision statistics, which have standard definitions defined by this matrix as

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}}, \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (2)$$

2.3. Data Acquisition

All images were obtained in FITS format (Pence, 1999) from the publicly available Big Bear Solar Observatory (BBSO) FTP archive (<http://www.bbso.njit.edu/>). We downloaded

a matching list of 523 images previously obtained and analyzed by the AAFDCC module developers, spanning almost two years from 15 July 2000 to 4 July 2002. We chose this date range to match previous work that compared to an earlier human-made list of filaments (Bernasconi, Rust, and Hakim, 2005; Pevtsov, Balasubramaniam, and Rogers, 2003). Although these papers were both focused on determining chirality of filaments, rather than just detection, the efforts provided additional scrutiny over these images and their identified filaments. Unfortunately, Pevtsov, Balasubramaniam, and Rogers (2003) did not record the subset of images used during this time frame, and therefore we cannot easily make a direct comparison to it. Although not further analyzed in our study, we note that the comparative work of Bernasconi, Rust, and Hakim (2005) showed superior results for the FFT module over the traditional human-based identification of Pevtsov, Balasubramaniam, and Rogers (2003). We also note several algorithms for filament detection have been developed in recent years (Gao, Wang, and Zhou, 2002; Shih and Kowalski, 2003; Qu *et al.*, 2005; Fuller, Aboudarham, and Bentley, 2005), but we chose to use one specific module with a well known and available dataset for our preliminary investigations.

After establishing the image dataset, we received the corresponding event metadata produced by the AAFDCC module. The module has been developed to produce several types of metadata output, including: a text file containing attributes of each discovered filament, a binary bitmap mask for each image showing each filaments' location and area on the solar disk, and a VOEvent file to submit to the heliophysics event knowledgebase (HEK) for each filament event detected. We chose to use the text-based filament attribute results files during development as they offer more detailed information than the HEK events, including many attributes that describe the rough physical characteristics of each filament.

The filament data is in the form of tabular ASCII files (one file for each image), with each row representing a filament detected by the module. We focus on the Cartesian coordinate (x, y) offset of filament center from sun center, average filament angle (with respect to the horizontal equator), and length of filament (in thousandths of solar radii), as these attributes are used to create our labeled datasets. We also require the use of several FITS header attributes (metadata attached to each image) to normalize each image to the same dimensions as the AAFDCC module reports on – namely: the solar disk center (in pixel coordinates) and the solar disk width.

2.4. Data Verification

Before processing the above mentioned data attributes, we first have a look at the collection of values present in the dataset. This serves as a data integrity verification, and a personal sanity check, so that we can be more confident when relying on this module to derive labels.

Figure 3 shows four filament attributes provided by the AAFDCC module. First, notice the expected distributions of filament center pixel locations (as offsets of solar disk center). Since the BBSO images are 2048×2048 pixels, we can see the offsets from center appropriately range between ± 1024 pixels. Notice the hot spots of filament activity seen in the y -axis ranges representing the solar active region belt at that time, and the evenly distributed x -axis values – both of which confirm the common assumptions of filament occurrence.

The filament length is displayed in a log-based scale, as the distribution has a very long right tail – the non-transformed values (in thousandths of solar radii) have a minimum of 19, maximum of 1195, and median of 80. The distribution of the filament lengths appears like a power law with negative coefficient in the logarithm of the filament length, a new result as far as we know, with perhaps significant implications for models of the formation of filaments (Martens and Zwaan, 2001). Finally, notice the bi-modal distribution of filament

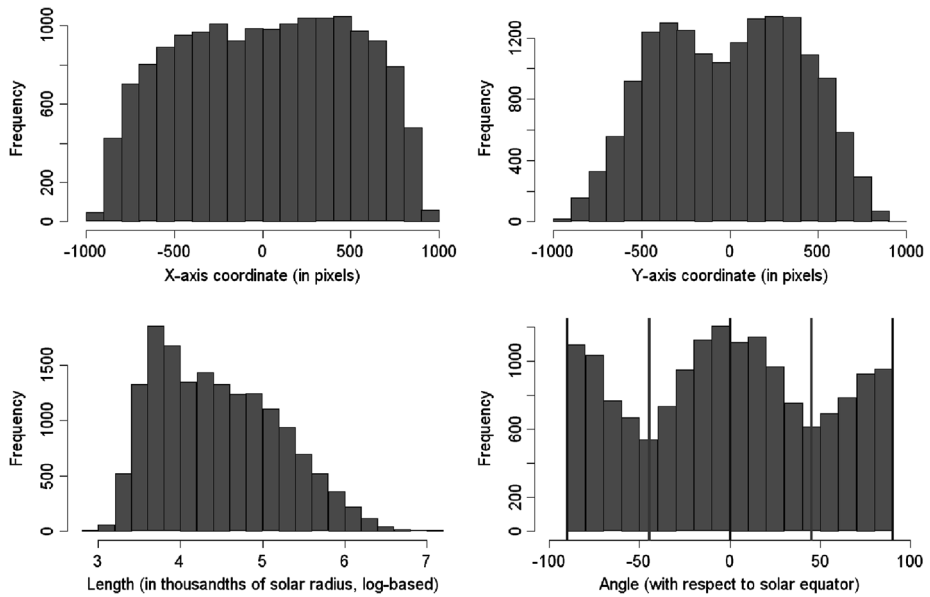


Figure 3 Analysis of selected attributes from the AAFDCC module metadata.

angle average, which conforms to the valid range of $\pm 90^\circ$, and seems to indicate a strong preference towards horizontal and vertical filament inclination. This interesting finding may also constitute a significant constraint on models for the evolution of solar filaments, but a more thorough investigation is required to relate the results to prior work (Martin and Alexander, 2009). As we will discuss in the next section, the worst case for our label creation is 45° angles, which results in the largest and noisiest rectangular boxes. However, we can clearly see the lowest frequency of values occurs near $\pm 45^\circ$, indicated by the vertical red lines, and this largely mitigates problem.

We also analyze the necessary FITS header attributes in a similar manner, shown in Figure 4. The top two charts display the histogram of values given for the solar disk center (x and y pixel coordinates), with a red line inserted at the absolute image center (1024 pixels). Both attributes appear normally distributed, but we can see that the Sun center is on average slightly lower left from the actual image center. A side-by-side boxplot of center coordinate values is also shown, and indicates very similar distributions. Finally, we present a histogram of solar disk width. The distribution is bi-modal with highest frequencies at the extreme values, which is expected due to yearly cycles as the Earth travels around the Sun in an elliptical path.

3. Data Preprocessing

We perform several data transformation steps on the images and events. First, the image files were uncompressed and converted to TIFF format. The advantage of TIFF is direct pixel-based image data which can easily be accessed and manipulated by standard computer vision techniques. It is also a more compact (but still lossless) file format than FITS, resulting in identical images of about half the file size.

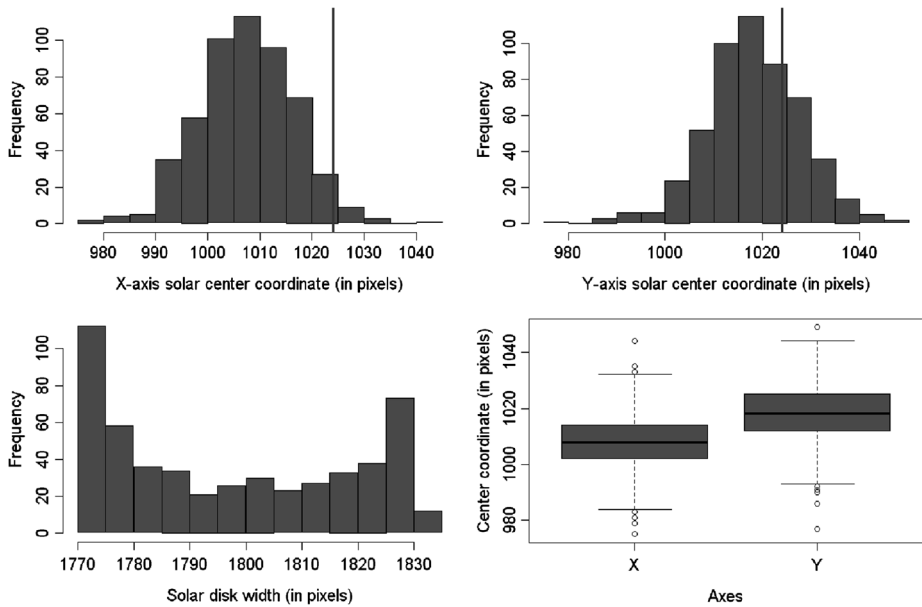


Figure 4 Analysis of values found in FITS header keywords.

Table 2 The total number and size of cells created for the three grid sizes.

Grid size (in cells)	Total cells (per image)	Cell size (in pixels)
16 × 16	256	128 × 128
32 × 32	1024	64 × 64
64 × 64	4096	32 × 32

3.1. Labeling Methods

The most important step in our process is transforming the raw images and metadata into usable data for our TFR module. This involves segmenting the images by a rectangular grid into smaller pieces, or *image cells*, and defining a *class label* for each cell indicating whether it is a filament or non-filament. By varying the number of rows and columns (equally) in the segmentation grid, we established three levels of granularity: 16 × 16, 32 × 32, and 64 × 64 cells per image, shown in Table 2. For each grid size, the parameter values are extracted from each image cell and stored individually as a *parameter vector* containing the ten numerical values from Table 1. We then add a class label to each parameter vector based on whether or not the cell is within a filament region determined by the AAFDCC module’s metadata.

Imposing a fixed grid structure is a common approach to discretizing raster-based data into manageable vector-based objects. It is also well known that this process is inherently noisy because of the data-independent grid placement (Shekhar and Chawla, 2003). In our case perhaps only a portion of an image cell shows presence of a filament, or vice versa. No matter which label is assigned to the cell, it will be partially incorrect, and training/testing will suffer. We note that this problem will always be present because of the blob-shaped events overlaying a fixed grid of cells. So while more exact event labels (such as bitmap

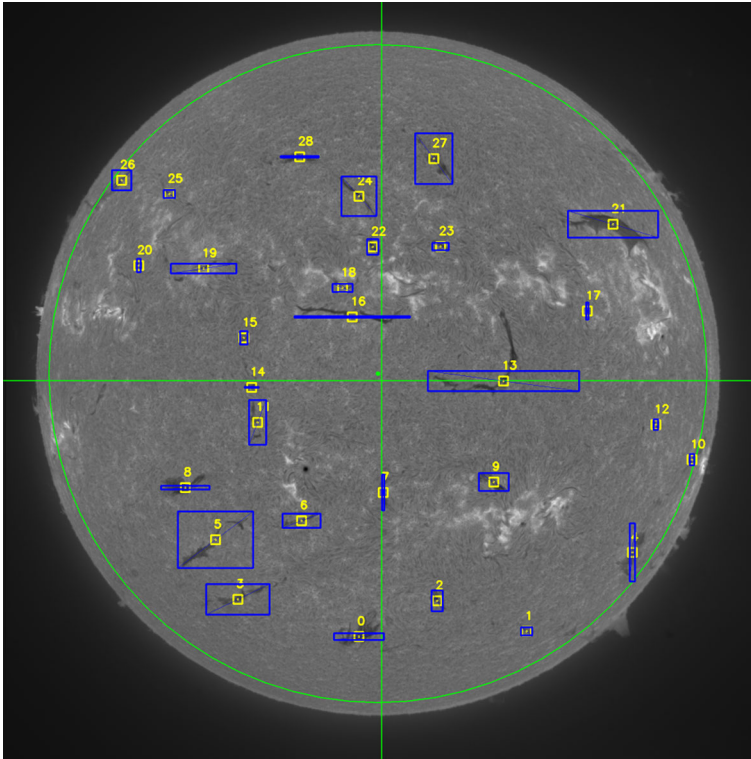


Figure 5 An example BBSO H α image with overlaid data attributes used to create three labeling methods. Each filament has a small *center* labeled region (yellow box), and a larger *est-MBR* region (blue box) based on the filament size and angle. The third label method (*sub-MBR*) uses a combination of these two regions, by removing all the *est-MBR* filament cells which are not *center* filament cells to reduce the noise from these boundary regions.

masks) might mitigate (but not eliminate) these issues, our initial investigation focuses on general event descriptions that can be generalized across many modules and event types.

We explore three labeling methods based on the filaments' centers, an estimated minimum bounding rectangle (MBR) which roughly contains the blob-shaped filament, or a combination of the two, respectively named: *center*, *est-MBR*, and *sub-MBR*. Figure 5 shows an example of these labeling methods which we will discuss in detail in the following subsections.

3.1.1. Label: Center

The *center* method is the simplest labeling approach based on only the center coordinates of each filament, derived from the AAFDCC module's solar disk center position, and the center offsets given for each filament. We create square regions 1/10th the width and height of a cell around each filament center, shown in yellow in Figure 5.

Rather than just marking cells that contain the exact center points, we use sufficiently small boxes for better representation of filaments whose centers were too close to the cell edge. Through empirical observations we found these neighboring cells to also predominantly contain filaments. Therefore, by labeling them all in this manner it not only helped

reduce noise between classes, but it also allowed us to gather more cells for the much more scarce filament class label.

3.1.2. Label: *Est-MBR*

The second method is our estimated minimum bounding rectangle (MBR), *est-MBR*, which roughly places a rectangle box around each filament, shown as blue rectangles in Figure 5. The box dimensions are derived from the length and angle of each filament with respect to the solar disk size and assumed horizontal axis. The angle measurement, however, is an average angle over the entire filament and may not always extrapolate well when applied to the filaments total length. Also, if the angle is close to horizontal or vertical, the derived box will be quite narrow. For example, consider a highly curved filament, whose angle may average to near horizontal, but whose length (because of curvature) extends the box far beyond than the filament's actual area, such as #13 in Figure 5.

Furthermore, notice that *est-MBR* labeling can contain many cells that should be labeled non-filament because the natural blob-shaped filaments do not translate well to rectangular regions. Similarly, we notice that *center* labels contain many should-be-filament-labeled cells – especially with smaller cell sizes where the center captures less of the actual filament. It should be clear that even with exact filament MBRs the transformation to grid-based cells will inevitably produce some incorrectly labeled cells.

3.1.3. Label: *Sub-MBR*

Our final labeling method, *sub-MBR*, attempts to minimize the number of noisy cells by first labeling the center cells as filaments, and then removing (“subtracting”) the rest of the *est-MBR* cells from the dataset. It should be clear that almost all cells within the *center* regions are filaments, and that most of the cells outside of *est-MBR* are non-filaments (again #13 in Figure 5 is a counter example). Therefore, we use the *est-MBR* label as an indicator of (at least a part of) a region of cells that may contain erroneous labels. By simply discarding these cells that have a higher likelihood of being labeled incorrectly, we expect to reduce ambiguity between class labels and thereby increase classifier accuracy.

3.2. Dataset Creation

Regardless of labeling method we finish by labeling all cells “remove” which are outside of a 75° heliocentric area of interest, shown as a circular ring in green on the edge of the solar disk in Figure 5. This is done to match the preprocessing step taken by the AAFDCC module, with a justification that the edges of the solar disk are too distorted for proper chirality determination – and technically, filaments are not observed (instead called prominences) beyond the solar disk (Bernasconi, Rust, and Hakim, 2005). We chose to perform this procedure last to catch bounding boxes which may have erroneously extended beyond the area of interest. Figure 6 shows the difference in cell labels over all nine variations of the dataset for a magnified region of interest from the previously shown full-disk image in Figure 5 – with the region's original metadata shown at scale on the top for reference.

Finally, we save each cell's parameter vector as a data instance in ARFF format (defined at: <http://www.cs.waikato.ac.nz/ml/weka/arff.html>) used by the popular and open source machine-learning software WEKA (Hall *et al.*, 2009). All cells labeled “remove” are now discarded entirely, leaving a binary class label of either *filament* or *non-filament*. Because of proper data randomization during experiments, we include two identification

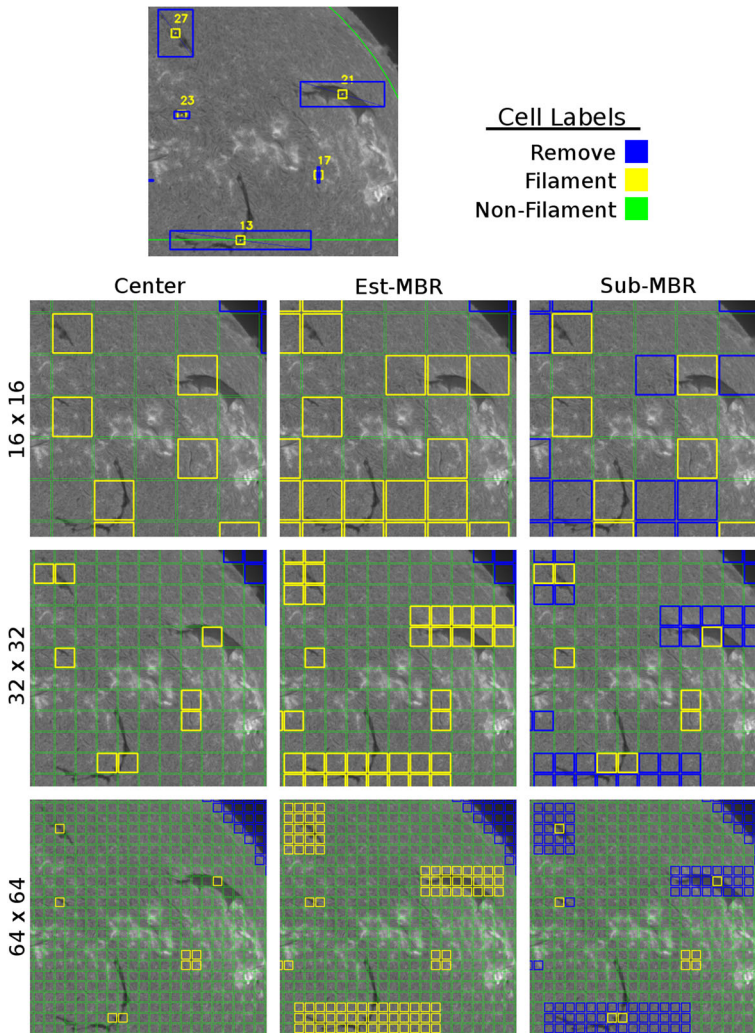


Figure 6 An example of all possible cell labels used to train the classifiers. The region of interest is shown to scale at the top, displaying the metadata used to derive the labels. The grid size increases from top to bottom and label methods are from left to right as *center*, *est-MBR*, and *sub-MBR*. The cell labels' class is indicated by its outline color: filament (yellow), non-filament (green), and remove-from-dataset (blue).

attributes for each cell: file name and cell index, so that we can trace each cell back to its proper place in a specific image. These values are placed in the front of the parameter vector, whereas the label is traditionally placed at the end. For example, consider the cell highlighted in Figure 1, and assuming it contains a filament, we have the parameter vector $cell_{4,3} = \langle \text{image-name}, 50, 4.4309, 33.8157, 13.9262, 1.2423, 5.1317, 31.1156, 0.0645, 0.0002, 2.9546, 0.0014, \text{filament} \rangle$.

The identification attributes are removed prior to training and testing so that they do not influence any outcomes, such as learning that a certain cell index is prone to filaments. Figure 7 shows the distribution of data instances for each variation of our dataset.

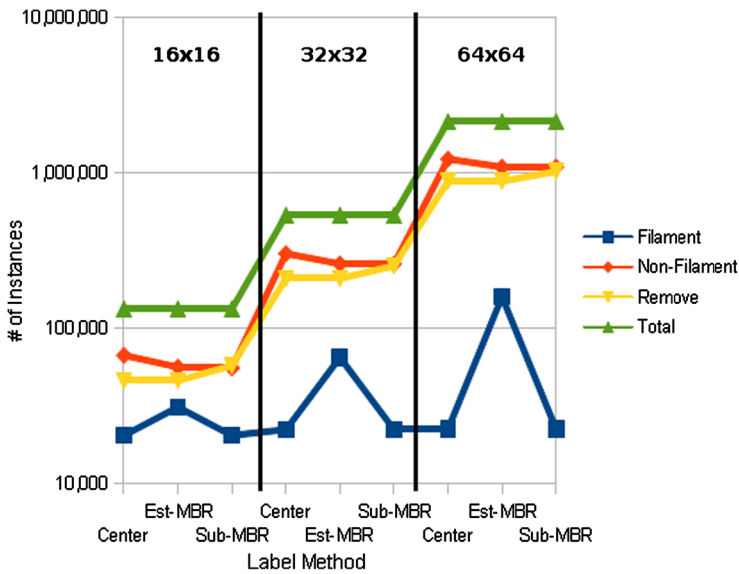


Figure 7 The total number of each type of cell label prior to final dataset creation. The label types are shown over all three grids and label methods, including total number of cells.

4. Experiments and Results

4.1. Overview

The focus of our experiments is to determine how well our trainable feature recognition (TFR) module can learn to detect solar filaments, based on metadata created by the Advanced Automated Solar Filament Detection and Characterization Code (AAFDCC) module (Bernasconi, Rust, and Hakim, 2005). The TFR module is part of a larger framework previously developed by the SDO FFT at Montana State University to create a general purpose CBIR system for solar image archives (Banda, 2011). Before describing the setup, we first mention several important assertions.

The AAFDCC module has been extensively developed over ten years for the specific purpose of detecting and analyzing filaments in $H\alpha$, and we do not expect to – or intend to – out-perform it with our general purpose CBIR framework and its recent adaptation for these experiments. Note that we focus only on the detection of filaments in single images, and we do not perform filament-specific analysis (such as determining chirality), or attempt to track filaments over time or merge broken filament segments – all of which the AAFDCC module now does. We do, however, still gain valuable information as regards our ability to detect filaments while also independently analyzing and validating the automatically reported results of the AAFDCC module.

It is important to reiterate that we have no equivocally true labels to verify whether or not an event (in this case a filament) actually exists in a specified region of an image. This is an inherent problem with solar physics imagery, as even human-based labels can contain biased results (Bernasconi, Rust, and Hakim, 2005). Therefore, we cannot simply evaluate accuracy objectively, *e.g.*, the module reported 92 % of all existing filaments in the dataset, because the true set of filaments is unknown. Instead, we have to assume truth in the AAFDCC module-based labels that are used to train and evaluate our TFR module. This means that,

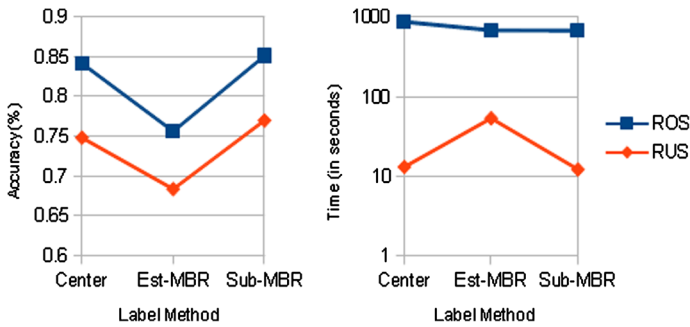


Figure 8 The accuracy and timing (in seconds) results of random over-sampling (ROS) and random under-sampling (RUS) methods for dataset class balancing.

for example, if we reported 100 % detection accuracy, then what we are really saying is that we recognized exactly the filaments reported by the AAFDCC module, and no others – even if others exist, or some do not.

4.2. Class Balancing

There are nine total datasets created by the combination of three grid sizes and three label methods, and each experimental run starts with one of these initial datasets. We first randomized the data and balanced the number of data instances in each class to avoid a biased classification model, as there are far fewer cells labeled as filament than non-filament. With highly skewed class sizes, the classifier may find that the best training strategy simply ignores the smaller class. For example, if 90 % of the data was labeled non-filament, then a classifier could simply say *everything* has a non-filament label, and achieve 90 % accuracy. Obviously this level of accuracy is a false assurance of our prediction abilities. Therefore, we explored the typical choices of random over-sampling (ROS) and random under-sampling (RUS) to balance the dataset (Japkowicz, 2000), where ROS randomly duplicates instances of the rarer (less frequent) class to match the number of instances in the abundant (more frequent) class, and RUS randomly removes instances from the more frequent class until it matches the size of the less frequent class.

While over-sampling was generally capable of producing better results, it was more so because of repetitive training on identical cells, some copies of which could have been present in both train and test sets, than simply having a larger dataset. Also, under-sampling produces the smallest dataset possible, while retaining the important – and fully unique – set of filament cells, which drastically decreases memory requirements and run times during experimentation. Figure 8 shows the classification accuracy percentage and running time in seconds (to build the model and evaluate it) for all four classifiers and all three label types on the 32×32 grid. Regarding memory requirements, the 32×32 grid with the center label method, for example, produced over 12 times more non-filament cells than filament cells (roughly 300 000 vs. 25 000). While this results in an over-sampled dataset of 600 000 instances, the under-sampled set is only 50 000 instances, or about 8 % of the size of the over-sampled set and less than one-sixth (16.67 %) the size of the original (unbalanced) dataset. The final dataset sizes (after RUS) are given in Table 3, where each class is exactly half of the total instances reported (*i.e.*, balanced).

Table 3 Total data instances for each dataset variation after all preprocessing.

Label	16 × 16	32 × 32	64 × 64
<i>Center</i>	41 038	44 760	45 080
<i>Est-MBR</i>	62 426	130 278	320 478
<i>Sub-MBR</i>	41 038	44 760	45 080

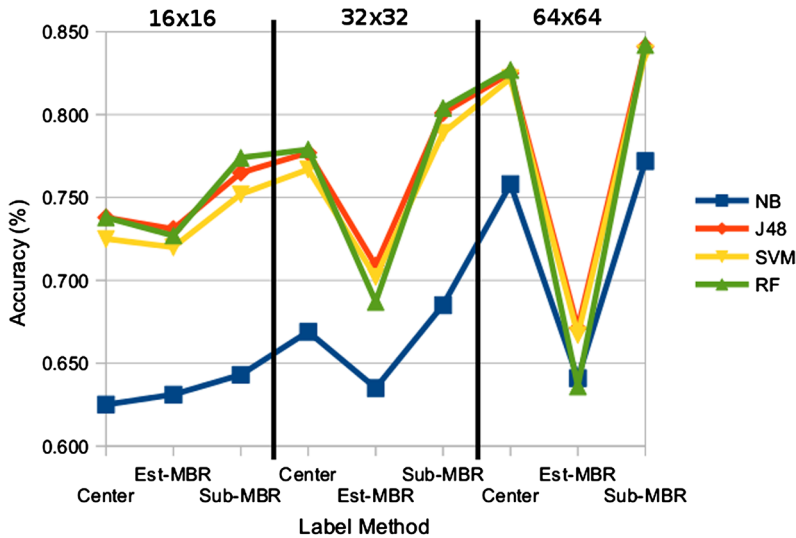


Figure 9 The classifier average accuracy results for each type of experimental run.

4.3. Comparative Evaluation

We evaluate all nine datasets over four different classification models to analyze the 36 unique combinations of grid sizes, label methods, and classifiers. Since this is an overview of how well each combination might work, and not how we can best fine-tune a classifier for a specific grid or label method, we use the default settings (provided by WEKA) for the following four classifiers: NB, J48 (WEKA’s version of the C4.5 decision tree), RF, and SVM. While many other options exist, these four are quite commonly used by the machine-learning community.

Each unique combination was evaluated 12 times and the results were averaged for statistical reporting. Each evaluation began with an initial dataset, before being balanced (RUS), and then randomly separated into a standard two-thirds training set and one-third testing set. Randomization occurs during both operations (for each new run) to ensure there is no bias in the order or selection of non-filament cell instances in the dataset.

We record the statistics: accuracy, TP, FP, FN, TN, run time (in seconds), precision, and the ROC (receiver operating characteristic) curve – all of which can be found in full detail in the [Appendix](#).

First we present the average accuracy of each unique combination, shown in Figure 9. In general, we can see that the *est-MBR* label performs the worst over each grid size, and the *sub-MBR* label performs slightly better than the *center*, which was expected. While Naïve Bayes is less accurate overall, the other three classifiers perform similarly, with Random Forests being the most accurate in most cases. The best accuracy of 84.2 % was achieved

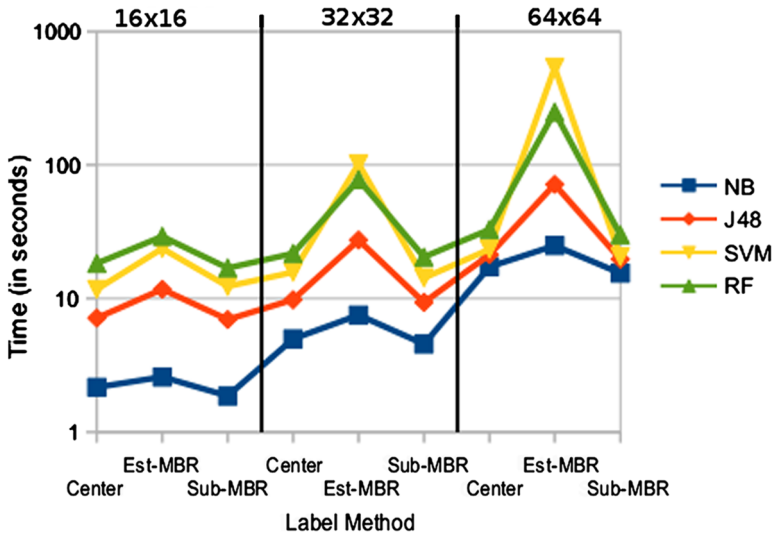


Figure 10 Run time (log seconds) for each classifier over grid size and label type.

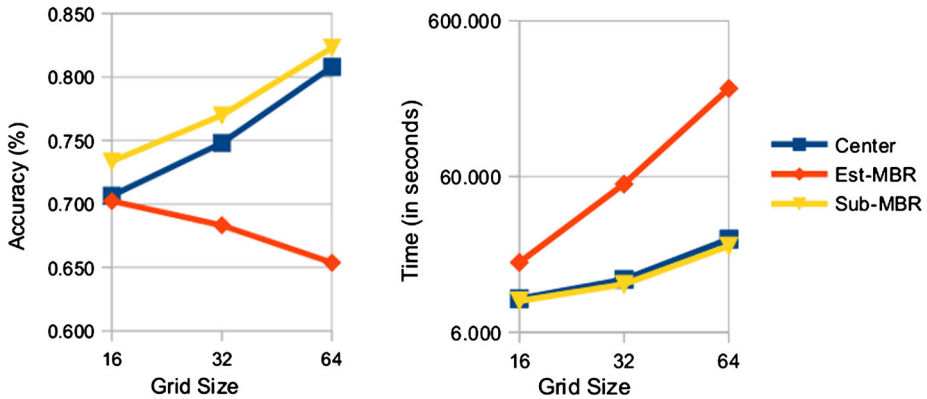


Figure 11 The average accuracy and time (in seconds) results of all four classifiers aggregated over each type of grid and label.

on the 64x64 grid using the *sub-MBR* labels and the RF classifier. Also, a noteworthy result of 82.7 % was recorded in the same scenario but while using the *center* labels. In most cases, the J48 classifier was very closely behind RF. Again, we recall that “accuracy” here refers to an agreement with the AAFDCC module metadata, and not an absolute indisputable accuracy.

Figure 10 presents the time taken (in seconds) to train and test each classifier. The time scale is logarithmic, and it is clear that the *est-MBR* labels take significantly more time to evaluate. Not surprisingly, the classifiers took much longer on the larger datasets created from finer grid segmentation. Note that the fastest classifier in all cases was NB, and the slowest was RF or SVM.

The general effects of grid size and label method are better shown when we aggregate the accuracy and run time of the classifiers together. Presented in Figure 11, we can see

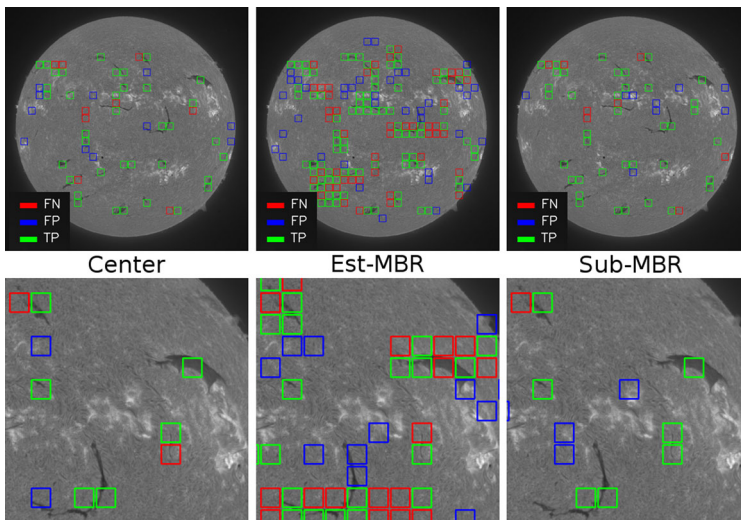


Figure 12 An example of visualizing the confusion matrix results on an image.

that as the grid size increases, *est-MBR* accuracy declines while the other two label methods increase similarly in accuracy. This makes sense, because there will be more cells labeled incorrectly and subsequently trained on with the *est-MBR* method. We also note that the difference of performance between the *center* and *sub-MBR* methods shrinks as we increase the grid size, from: 2.7 % to 1.5 %.

4.4. Visualizing the Results

While the previous experiment gave us an excellent indication of our ability to detect filaments, we now visually explore how well we actually distinguish filaments in the segmented images. We display each tested image and outline cells for three outcomes:

- i) The classifier labeled the cell filament, and so did the AAFDCC module.
- ii) The classifier labeled the cell a filament, but the AAFDCC module did not.
- iii) The classifier did not label the cell as a filament, but the AAFDCC module did.

These three outcomes directly correspond to the confusion matrix results of TP, FP, and FN, respectively. The TN cells – the classifier and AAFDCC module agree the cell does not contain a filament – were ignored here for a cleaner visualization, and because many of these should-be-tested cells are missing from an image (due to under-sampling), which would make the results less clear.

An example of our visualization is shown in Figure 12 for the RF classifier on the 32×32 grid. The same image and region of interest from Figure 6 are shown again, with the three columns representing each labeling methods. We can see that in all three cases the majority of cells are correctly classified (green – TP), but the *est-MBR* labels have many more incorrectly labeled cells (red – FN, and blue – FP) almost all of which are in ambiguous and noisy areas of filament regions. A great example of the power of a supervised classifier can be seen for the filament in the bottom left of the cut-out region (again #13 from

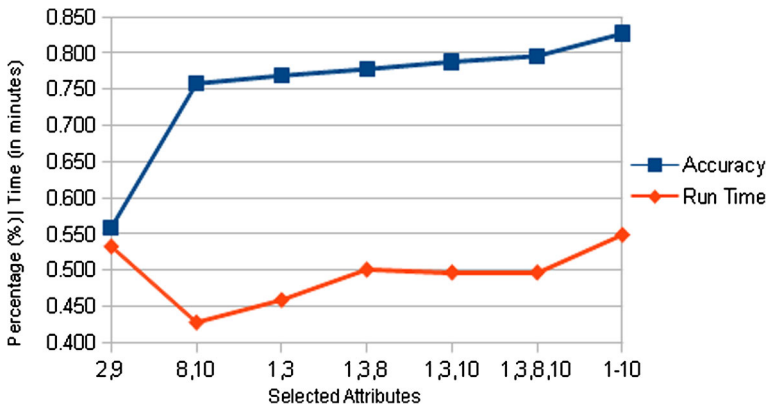


Figure 13 The classification accuracy and time (in seconds) over several parameter subsets.

Figure 5). Recall the poor *est-MBR* label of this filament (Figure 6), and notice the results of testing our classifier. We can see that two cells are labeled FN because our label says those are filament cells, even though they are clearly not. Similarly, two cells are labeled FP because our labels says those are non-filament cells, even though they clearly are. Note that all four of these cells negatively impact our current measure of classification accuracy.

4.5. Attribute Evaluation

It is important to investigate which attributes (parameters) work best for different kinds of solar events as well as different types of solar imagery. If we can reduce the attributes needed for each classifier, then we can better streamline our TFR module when tackling many simultaneous events at once in the future.

Evaluation was performed identical to the full comparative evaluation, except we chose a different subset of attributes instead of varying the grid size and label method. All evaluations were performed on the 64×64 grid with the *center* labeling method. We first re-run the evaluation with all attributes as our initial benchmark, and then we try each attribute individually, before trying several promising subsets. The attribute subsets were chosen through empirical evidence from their individual results as well as through several attribute evaluation ranking methods provided by WEKA, which list the attributes in order of importance towards distinguishing the two classes. These methods included: Information Gain, CfsSubsetEval, and Principal Component Analysis. Note that all of these methods allowed us to chose attributes independently of any domain expertise or knowledge of which attributes might be more promising. While expert knowledge could potentially help these decision processes, we focus on the automated/non-expert solution guided by the machine-learning community.

The accuracy results for these subsets are presented in Figure 13. Included is a subset of poorly performing attributes, showing that even when combined, parameters 2 and 9 do not work well in this classification task. However, we find a high level of accuracy can still be achieved when using only four parameters (1, 3, 8, and 10). This elimination of 60 % of the

parameters results in an accuracy of 79.6 %, compared to the original 82.7 %, a reduction of only 3.1 %.

5. Conclusion and Future Work

This work is the first attempt to establish a comparative evaluation environment for automated event detection modules. We have shown validation of the AAFDCC module (Bernasconi, Rust, and Hakim, 2005) over a two year span of images. Furthermore, our TFR module successfully used this metadata to achieve a classification accuracy of over 82 %, using only *center* labels and a single (J48) decision tree classifier with no algorithmic tuning or domain expertise. We also find that we can achieve a quite similar accuracy using only a handful of the original attributes, meaning more computationally efficient training and detection of filaments within our module. The success of this work provides the motivation to fully realize a comparative evaluation framework for automated event detection modules based on our SDO FFT TFR module.

The next step involves extending our work to larger datasets and other events. We plan to first re-evaluate our filament detection abilities with the AAFDCC module's HEK reported metadata over the entire 10 years of module operation for a fully comprehensive look at solar filaments in the automated reporting age. After that we will move on to other SDO FFT modules and events, such as active regions, coronal holes, and sunspots. In preparation for this future work, we recently published an initial six month dataset containing over 15 000 SDO images and 24 000 region-based event labels from six automated SDO FFT modules (Schuh *et al.*, 2013b).

We also plan to approach open research topics in spatial/temporal classification and evaluation, primarily motivated by our ongoing discussion of filament #13 throughout this paper. By incorporating spatial knowledge of image cell neighborhoods, we could further mitigate the negative effects of erroneous labeling. For example, the penalty for mis-classifying a cell that has neighbors of each class type could be less than that of mis-classifying a cell amongst neighbors of all one type. In other words, if we erroneously label a single cell a filament in the middle of a quiet sun (non-filament) region, this would be a "worse" error than if the cell happened to be adjacent to an actual filament. Furthermore, this spatial knowledge can be included in the classifier itself (especially given complementary domain knowledge), weighting its prediction decision before we even evaluate the outcome. For example, it would be nearly impossible for a single cell to be an active region if all neighbor cells were labeled as a coronal hole. Both of these applications contribute to active areas of research in computer science.

Acknowledgements This research and development project was supported by two NASA Grant Awards: No. NNX09AB03G, funded from the NNH08ZDA001N-SDOSC solicitation, and No. NNX11AM13A, funded from the NNH11ZHA003C solicitation. We would also like to thank our internal reviewers as well as the Big Bear Solar Observatory/New Jersey Institute of Technology and the Global High Resolution H α Network for providing and maintaining the ftp image data archive.

Appendix

The appendix contains the complete tabulated results for the discussed experiments. The summarized charts and insights were gathered from this data.

Table 4 Complete attribute evaluation results. Note we do not try every possible combination of attributes, as discussed in the text. The first column denotes the set of attributes used for the evaluation.

Atts	Acc	TP	FN	FP	TN	Time	ROC	PRE
1	0.668	0.337	0.163	0.169	0.331	26.520	0.73	0.67
2	0.511	0.258	0.241	0.249	0.253	28.800	0.52	0.51
3	0.693	0.350	0.151	0.156	0.343	26.100	0.75	0.69
4	0.624	0.265	0.235	0.141	0.359	19.560	0.66	0.63
5	0.596	0.303	0.198	0.206	0.294	26.880	0.64	0.6
6	0.582	0.296	0.205	0.213	0.286	27.240	0.62	0.58
7	0.659	0.336	0.166	0.175	0.323	24.780	0.71	0.66
8	0.741	0.381	0.119	0.140	0.360	19.740	0.81	0.74
9	0.530	0.268	0.231	0.239	0.263	28.860	0.55	0.53
10	0.750	0.362	0.137	0.113	0.388	21.360	0.82	0.75
2, 9	0.559	0.303	0.196	0.245	0.255	31.980	0.58	0.56
1, 3	0.769	0.403	0.096	0.135	0.366	27.540	0.83	0.77
8, 10	0.758	0.391	0.109	0.133	0.367	25.680	0.82	0.76
1, 3, 8	0.778	0.410	0.090	0.133	0.368	30.060	0.84	0.78
1, 3, 10	0.788	0.414	0.085	0.127	0.374	29.820	0.85	0.79
1, 3, 8, 10	0.796	0.420	0.083	0.121	0.376	29.820	0.86	0.8
All	0.827	0.432	0.068	0.105	0.395	32.940	0.89	0.83

Table 5 Comparative evaluation results for all dataset variations. The three grid sizes are separated by a horizontal line, and presented in order of: 16×16 , 32×32 , and 64×64 . Time is given in seconds, and the receiver operating characteristic (ROC) and precision (PRE) are averaged over both classes.

Box	Model	Acc	TP	FN	FP	TN	Time	ROC	PRE
<i>Center</i>	NB	0.625	0.406	0.095	0.281	0.218	2.160	0.683	0.645
<i>Center</i>	J48	0.738	0.400	0.100	0.161	0.339	7.140	0.771	0.742
<i>Center</i>	SVM	0.725	0.367	0.132	0.143	0.358	11.700	0.725	0.725
<i>Center</i>	RF	0.738	0.402	0.096	0.166	0.336	18.300	0.798	0.743
<i>Est-MBR</i>	NB	0.631	0.405	0.094	0.275	0.225	2.580	0.688	0.651
<i>Est-MBR</i>	J48	0.731	0.372	0.127	0.142	0.359	11.700	0.773	0.731
<i>Est-MBR</i>	SVM	0.720	0.362	0.137	0.143	0.358	23.700	0.720	0.720
<i>Est-MBR</i>	RF	0.727	0.389	0.110	0.163	0.338	29.220	0.792	0.730
<i>Sub-MBR</i>	NB	0.643	0.408	0.091	0.266	0.235	1.860	0.708	0.664
<i>Sub-MBR</i>	J48	0.765	0.401	0.100	0.135	0.365	6.960	0.799	0.767
<i>Sub-MBR</i>	SVM	0.752	0.378	0.121	0.127	0.374	12.300	0.752	0.752
<i>Sub-MBR</i>	RF	0.774	0.413	0.086	0.140	0.361	16.920	0.841	0.777

Table 5 (Continued)

Box	Model	Acc	TP	FN	FP	TN	Time	ROC	PRE
<i>Center</i>	NB	0.669	0.393	0.107	0.223	0.276	4.980	0.736	0.679
<i>Center</i>	J48	0.777	0.411	0.087	0.136	0.365	9.780	0.824	0.780
<i>Center</i>	SVM	0.767	0.392	0.109	0.125	0.375	15.720	0.767	0.767
<i>Center</i>	RF	0.779	0.421	0.081	0.141	0.358	21.780	0.839	0.783
<i>Est-MBR</i>	NB	0.635	0.388	0.112	0.252	0.248	7.500	0.697	0.647
<i>Est-MBR</i>	J48	0.709	0.344	0.156	0.135	0.366	27.420	0.764	0.710
<i>Est-MBR</i>	SVM	0.702	0.340	0.159	0.138	0.362	101.760	0.702	0.703
<i>Est-MBR</i>	RF	0.687	0.366	0.134	0.178	0.321	77.640	0.751	0.689
<i>Sub-MBR</i>	NB	0.685	0.397	0.103	0.212	0.287	4.560	0.757	0.694
<i>Sub-MBR</i>	J48	0.801	0.414	0.085	0.114	0.387	9.360	0.849	0.803
<i>Sub-MBR</i>	SVM	0.789	0.398	0.102	0.109	0.391	14.280	0.789	0.789
<i>Sub-MBR</i>	RF	0.804	0.424	0.075	0.120	0.381	20.460	0.871	0.807
<i>Center</i>	NB	0.758	0.383	0.116	0.126	0.376	17.280	0.825	0.759
<i>Center</i>	J48	0.825	0.422	0.077	0.098	0.403	21.300	0.879	0.826
<i>Center</i>	SVM	0.822	0.415	0.084	0.094	0.407	23.400	0.822	0.822
<i>Center</i>	RF	0.827	0.432	0.068	0.105	0.395	32.940	0.889	0.829
<i>Est-MBR</i>	NB	0.641	0.297	0.204	0.155	0.345	24.960	0.694	0.643
<i>Est-MBR</i>	J48	0.671	0.288	0.212	0.117	0.383	71.520	0.727	0.678
<i>Est-MBR</i>	SVM	0.667	0.294	0.206	0.127	0.373	536.460	0.667	0.671
<i>Est-MBR</i>	RF	0.636	0.338	0.162	0.203	0.298	249.360	0.696	0.637
<i>Sub-MBR</i>	NB	0.772	0.384	0.116	0.112	0.388	15.480	0.840	0.772
<i>Sub-MBR</i>	J48	0.841	0.423	0.076	0.084	0.417	19.680	0.894	0.841
<i>Sub-MBR</i>	SVM	0.837	0.419	0.080	0.083	0.418	20.580	0.837	0.837
<i>Sub-MBR</i>	RF	0.842	0.434	0.064	0.093	0.409	30.000	0.907	0.844

References

- Banda, J.M.: 2011, Framework for creating large-scale content-based image retrieval system (CBIR) for solar data analysis. Ph.D. thesis, Montana State University.
- Banda, J.M., Angryk, R.A.: 2010a, An experimental evaluation of popular image parameters for monochromatic solar image categorization. In: Guesgen, H.W., Murray, R.C. (eds.) *The Proceedings of the 23rd Florida Artificial Intelligence Research Society Conference (FLAIRS)*, AAAI Press, Menlo Park, 380–385.
- Banda, J.M., Angryk, R.A.: 2010b, Selection of image parameters as the first step towards creating a CBIR system for the solar dynamics observatory. In: Zhang, J., Shen, C., Geers, G., Wu, Q. (eds.) *International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, IEEE Conference Publishing Services, Piscataway, 528–534.
- Banda, J.M., Schuh, M.A., Wylie, T., McInerney, P., Angryk, R.A.: 2014, When too similar is bad: A practical example of the solar dynamics observatory content-based image-retrieval system. In: Catania, B., Cerquitelli, T., Chiusano, S., Guerrini, G., Kämpf, M., Kemper, A., Novikov, B., Palpanas, T., Pokorný, J., Vakali, A. (eds.) *New Trends in Databases and Information Systems Advances in Intelligent Systems and Computing* **241**, Springer, Cham, 87–95.
- Bernasconi, P.N., Rust, D.M., Hakim, D.: 2005, Advanced automated solar filament detection and characterization code: Description, performance, and results. *Solar Phys.* **228**, 97–117.
- Breiman, L.: 2001, Random forests. *Mach. Learn.* **45**, 5–32.

- Domingo, V., Fleck, B., Poland, A.I.: 1995, The SoHO mission: An overview. *Solar Phys.* **162**, 1–37.
- Domingos, P., Pazzani, M.: 1997, On the optimality of the simple Bayesian classifier under zero-one loss. *Mach. Learn.* **29**, 103–130.
- Fuller, N., Aboudarham, J., Bentley, R.D.: 2005, Filament recognition and image cleaning on Meudon H α spectroheliograms. *Solar Phys.* **227**, 61–73.
- Gao, J., Wang, H., Zhou, M.: 2002, Development of an automatic filament disappearance detection system. *Solar Phys.* **205**, 93–103.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: 2009, The WEKA data mining software: An update. *SIGKDD Explor.* **11**, 10–18.
- Handy, B.N., Acton, L.W., Kankelborg, C.C., Wolfson, C.J., Akin, D.J., Bruner, M.E., Carvalho, R., Catura, R.C., Chevalier, R., Duncan, D.W., Edwards, C.G., Feinstein, C.N., Freeland, S.L., Friedlaender, F.M., Hoffmann, C.H., Hurlburt, N.E., Jurcevich, B.K., Katz, N.L., Kelly, G.A., Lemen, J.R., Levay, M., Lindgren, R.W., Mathur, D.P., Meyer, S.B., Morrison, S.J., Morrison, M.D., Nightingale, R.W., Pope, T.P., Rehse, R.A., Schrijver, C.J., Shine, R.A., Shing, L., Strong, K.T., Tarbell, T.D., Title, A.M., Torgerson, D.D., Golub, L., Bookbinder, J.A., Caldwell, D., Cheimets, P.N., Davis, W.N., Deluca, E.E., McMullen, R.A., Warren, H.P., Amato, D., Fisher, R., Maldonado, H., Parkinson, C.: 1999, The Transition Region And Coronal Explorer (TRACE). *Solar Phys.* **187**, 229–260.
- Japkowicz, N.: 2000, Learning from imbalanced data sets: A comparison of various strategies. In: Japkowicz, N. (ed.) *AAAI Technical Report WS-00-05*, AAAI Press, Menlo Park, 10–15.
- Martens, P.C.H., Zwaan, C.: 2001, Origin and evolution of filament-prominence systems. *Astrophys. J.* **558**, 872–887.
- Martens, P.C.H., Attrill, G.D.R., Davey, A.R., Engell, A., Farid, S., Grigis, P.C., Kasper, J., Korreck, K., Saar, S.H., Savcheva, A., Su, Y., Testa, P., Wills-Davey, M., Bernasconi, P.N., Raouafi, N.E., Delouille, V.A., Hochedez, J.F., Cirtain, J.W., DeForest, C.E., Angryk, R.A., Moortel, I., Wiegelmann, T., Georgoulis, M.K., McAteer, R.T.J., Timmons, R.P.: 2012, Computer vision for the solar dynamics observatory. *Solar Phys.* **275**, 79–113.
- Martin, S.F., Alexander, W.: 2009, An investigation of exceptions to the solar hemispheric patterns, in small telescopes and astronomical research. In: Genet, R.M., Johnson, J.M., Wallen, V. (eds.) *Proceedings of Galileo's Legacy, a Celebration of Small Telescopes and Astronomical Research Four Centuries Later*, Collins Foundation, San Francisco, 1–2.
- Pence, W.D.: 1999, CFITSIO, v2.0: A new full-featured data interface. In: Mehringer, D.M., Plante, R.L., Roberts, D.A. (eds.) *Astronomical Data Analysis Software and Systems, ASP Conf. Ser.* **172**, 287–489.
- Pesnell, W.D., Thompson, B.T., Chamberlin, P.C.: 2011, The Solar Dynamics Observatory (SDO). *Solar Phys.* **275**, 3–15.
- Pevtsov, A.A., Balasubramaniam, K.S., Rogers, J.W.: 2003, Chirality of chromospheric filaments. *Astrophys. J.* **595**, 500–505.
- Qu, M., Shih, F.Y., Jing, J., Wang, H.: 2005, Automatic solar filament detection using image processing techniques. *Solar Phys.* **228**, 119–135.
- Quinlan, J.R.: 1986, Induction of decision trees. *Mach. Learn.* **1**, 86–106.
- Schuh, M.A., Banda, J.M., Angryk, R.A., Martens, P.C.H.: 2013a, Introducing the first publicly available Content-Based Image-Retrieval system for the Solar Dynamics Observatory mission. 44th AAS-SPD Meeting, #100.97.
- Schuh, M.A., Angryk, R.A., Ganesan Pillai, K., Banda, J., Martens, P.C.H.: 2013b, A large scale solar image dataset with labeled event regions. In: *Proc. International Conference on Image Processing (ICIP)*, 4349–4353.
- Shekhar, S., Chawla, S.: 2003, *Spatial Databases: A Tour*, Prentice Hall, Englewood Cliffs, 45–80.
- Shih, F.Y., Kowalski, A.J.: 2003, Automatic extraction of filaments in H α solar images. *Solar Phys.* **218**, 99–122.
- Tamura, H., Mori, S., Yamawaki, T.: 1978, Texture features corresponding to visual perception. *IEEE Trans. Syst. Man Cybern.* **8**, 460–472.
- Vapnik, V.N.: 1995, *The Nature of Statistical Learning Theory*, Springer, London, 267–290.