Prog Artif Intell (2012) 1:215–222
DOI 10.1007/s13748-012-0020-z

REGULAR PAPER

Region selection and image classification methodology using a non-conformity measure

S. González · J. Vega · A. Pereira · I. Pastor

Received: 10 November 2011 / Accepted: 3 June 2012 / Published online: 20 June 2012

Abstract All the pixels of an image do not contain the same amount of information. Typically, the borders of an image contain less information than the centre. This paper introduces a methodology to locate the most relevant regions in images. A relevant region is a set of pixels that contains suitable information to recognise an image class versus the rest. To perform a detailed analysis of images, they are divided into regions and for each region a conformal predictor is trained. The values of the non-conformity measure are used, on the one hand, to hedge the classifier outputs with confidence and credibility measures and, on the other hand, to choose the most relevant regions. The combination of the best regions and their predictors (one per each class) is used to classify new images. The dimensionality of the original images is reduced to the dimension of the regions combination. This technique has been applied to the classification of images in the Thomson scattering diagnostic of a nuclear fusion device: the TJ-II stellarator. There are five different types of images. A database of more than 1,200 TJ-II Thomson scattering images has been analyzed.

Keywords Images · Non-conformity measure · Classification · Conformal predictors · Region · SVM

Mathematics Subject Classification 68U10 · 68T10 · 68T45 · 62H35 · 94A08

1 Introduction

Nuclear fusion is a promising source of clean energy. In fusion by magnetic confinement, the plasma is trapped inside a magnetic field created by a set of magnetic coils. To achieve fusion, the plasma is heated until it reaches thermonuclear conditions. Physical plasma properties (temperature, density and confinement time among others) are measured at any time by a wide range of plasma diagnostics. Plasma diagnostics include visible and infrared cameras as powerful tools for data analysis. However, the analysis, interpretation and validation of images are more complex than the same tasks performed with waveforms generated by other diagnostics.

This paper introduces a novel methodology for image processing and classification using a non-conformity measure [1]. The technique has been applied to the analysis of images generated in the Thomson scattering diagnostic [2,3] of the TJ-II fusion device [4]. The Thomson scattering images are 2-D spectra, with the horizontal and vertical axes displaying, respectively, scattered wavelength and position along a plasma chord, which allows determining both plasma density and plasma temperature profiles. In the TJ-II case, each image belongs to one of five possible classes. These classes correspond, respectively, to CCD camera background (BCK), measurement of stray light without plasma or in a collapsed discharge (STR), image during ECH phase (ECH), image during NBI phase (NBI) and image after reaching the cut-off density during ECH heating (CFF).

So far, different alternatives to solve the Thomson scattering image classification problem have been carried out: support vector machines (SVM) [5,6], neural networks [7] and conformal predictors [8,9]. In all cases, very high success rates were obtained in the classification process. Also, in all the previous works, to avoid the large sizes of the original images, they were reduced using a 2-D wavelet transform [10] (Haar family).

In an image, not all the pixels contain the same amount of information. Typically, the borders of an image contain less information than the centre. One of the contributions of this

S. González (✉) · J. Vega · A. Pereira · I. Pastor
Asociación EURATOM/CIEMAT para Fusión,
28040 Madrid, Spain
e-mail: sergio.gonzalez@ciemat.es

paper is just to identify image zones with relevant information to solve a classification problem. Both the identification of the most relevant areas and the classification of images are carried out using a non-conformity measure. All the mentioned works about the classification of Thomson scattering images have used the full images regardless of the different amount of information provided by the different regions.

As explained, the Thomson scattering diagnostic generates five different types of images. To solve this multi-class classification problem, the one-versus-rest approach can be used. This strategy consists of the creation of k (5 in this case) different classifiers to distinguish a class c from the other $k-1$ classes. It is possible that the features (image pixels) that allow the recognition of the class c differ from the features that allow the recognition of other class d. Therefore, the most suitable region for the recognition of a class c can be different from the best region to recognize the class d. The developed technique identifies each one of these k regions and uses the combination of the regions to classify new images.

This paper is structured as follows: Sect. 2 includes a review of the conformal predictors technique. The region selection and image classification methodology is described in Sect. 3. Section 4 contains the experimental results obtained using the Thomson scattering images. Finally, Sect. 5 includes a discussion of the methodology introduced.

2 Summary on conformal predictors

Conformal predictors [11] are predictive models that hedge the output labels with values of credibility and confidence. Confidence indicates how sure we are on the classification given by the predictor. On the other hand, credibility indicates the quality of the data on which we base our decision [12] (how representative is the training dataset to perform the classification of a test sample). Conformal predictors have been applied to wide variety of techniques, such as SVM [12], neural networks [13], nearest neighbour [14] or regression [15].

There are many variants of conformal predictors. In this paper, we use the inductive conformal predictors (ICPs). The main advantage of the ICPs versus the transductive conformal predictors is its computational efficiency: the bulk of the computations is performed only once no matter the number of the samples to classify [1].

Let's assume a training set, $S = \{x_1^p, \ldots, x_m^p, x_1^c, \ldots, x_n^c\}$, composed by the union of a proper training set, $S^p = \{x_1^p, \ldots, x_m^p\}$, and a calibration set, $S^c = \{x_1^c, \ldots, x_n^c\}$. The proper training set is used to compute a predictive model. The calibration set, together with new samples, is used to compute the conformal measures (confidence and credibility). In the on-line version of the ICPs algorithm, the new samples classified by the predictors are added to the calibration set, together with their predicted labels. In contrast, in the off-line version, the new classified samples are not included in the calibration set and it remains unchanged [12].

Different techniques can be applied in order to build a model using the proper training set (e.g. neural networks, nearest neighbour, etc). In this paper, SVM [12] is used to build the models. In the case of a binary classification problem, SVM generates a separating hyperplane H that divides the feature space into two different regions (one per each class).

The conformal measures are computed using a non-conformity measure (a measure of the supportiveness of a sample). For every sample in the calibration set, its non-conformity value (α) is computed. The non-conformity measure used in this paper is:

$$\alpha_x = \begin{cases} -|\text{dist}(x, H)|, & \text{if the sample is properly classified} \\ |\text{dist}(x, H)|, & \text{if the sample is misclassified} \end{cases} \tag{2.1}$$

On classifying a new sample, the non-conformity values themselves do not characterise how different this sample is for a given calibration set S^C. It is necessary to compare the α value computed for the new sample with the α values of the samples in S^C. This value is called p_{value} and it is given by [11]:

$$p_{\text{value}} = \frac{\#\{i : \alpha_i \geq \alpha_{\text{new}}\}}{n+1} \tag{2.2}$$

where n is the number of samples in S^C.

In order to obtain the credibility and confidence values for a new sample, it is necessary to compute a p_{value} per each different class in the problem: assuming that the new sample belongs to one class x, a new α_x value and the corresponding $p_{\text{value\_}x}$ is computed. For a n-class classification problem, n different p_{values} are computed.

The values of confidence and credibility for a new sample are computed as [11]:

$$\text{Credibility} = p_1 \tag{2.3}$$

$$\text{Confidence} = 1 - p_2 \tag{2.4}$$

where p_1 is the largest p_{value} and p_2 is the second largest p_{value} computed for that sample.

3 Region selection and image classification methodology

Firstly, it is important to define what an image region is. In this paper, a region is a set of connected pixels in the image. The minimum size of a region is 1×1 (1 pixel) and the maximum size of a region is the full image. A single image contains several different regions. Some of them can be overlapped, and thus, an image pixel can belong to many regions.

The goal of the region selection and image classification (RSIC) methodology is to locate the most relevant regions of a set of images in a classification problem. These regions are the ones that allow a predictor to properly distinguish between different classes. In multi-class problems, it is possible that the region that allows the recognition of a specific class could not be the same than the one that allows the recognition of another class. Therefore, different image regions could be necessary in a multi-class classification problem.

The RSIC methodology is made up of five different steps: image division, models building, region evaluation, region selection and, finally, image classification. The following subsections describe these steps.

3.1 Image division

The image division is the first step of the RSIC methodology. In this step, images are divided into regions. Figure 1 shows an example of the division of images into four regions.

Expert knowledge can be used to identify the most accurate and efficient image division according to image characteristics. These regions contain the most relevant a priori information and help the system to select the most suitable ones to better identify each class.

3.2 Models building

In the previous step of the RSIC methodology, images have been divided into regions. Each one of these regions is a completely independent system and must be able to classify an image using only the information of the pixels contained in the region.

One versus rest SVM (OVR-SVM) has been chosen to solve the multi-class classification problem. In a k-class problem, k binary SVM classifiers are built. Each classifier generates a separation hyperplane between a class and the other $k-1$ classes. A new sample is assigned to the class for which the positive distance from the hyperplane is maximal [16].

In this case, one OVR-SVM system (containing k binary SVM classifiers) is trained per each image region. Each OVR-SVM system uses only the information in the region where it has been trained. To classify a new image, it uses only that region. The same proper training set (S^P) and calibration set (S^C) are used to train all the binary classifiers in all the OVR-SVM.

Figure 1 shows an example of a 3-class classification problem where the images have been divided into four regions. Each image colour represents one different class (each class contains one or more images). 4 OVR-SVM systems are trained and a total of $4 \times 3 = 12$ binary SVM classifiers are needed (3 binary SVM classifiers per each image region to classify one class versus the rest).

3.3 Region evaluation

After the generation of models on each image's region, it is necessary to select those ones that better distinguish between classes. Since it is a multi-class classifier, it is possible that a region that properly identifies a class does not work for another class. Therefore, the quality of a region is computed for each class separately (e.g. in the previous example, four quality measures are obtained for each region, one per each class).

The quality of a region r in the classification of a given class t is computed using the non-conformity values (α) (Eq. 2.1) of the calibration set (S^C) corresponding to the binary classifier that distinguishes the class t from the rest of the classes (α^{rt}). Using the non-conformity measure in

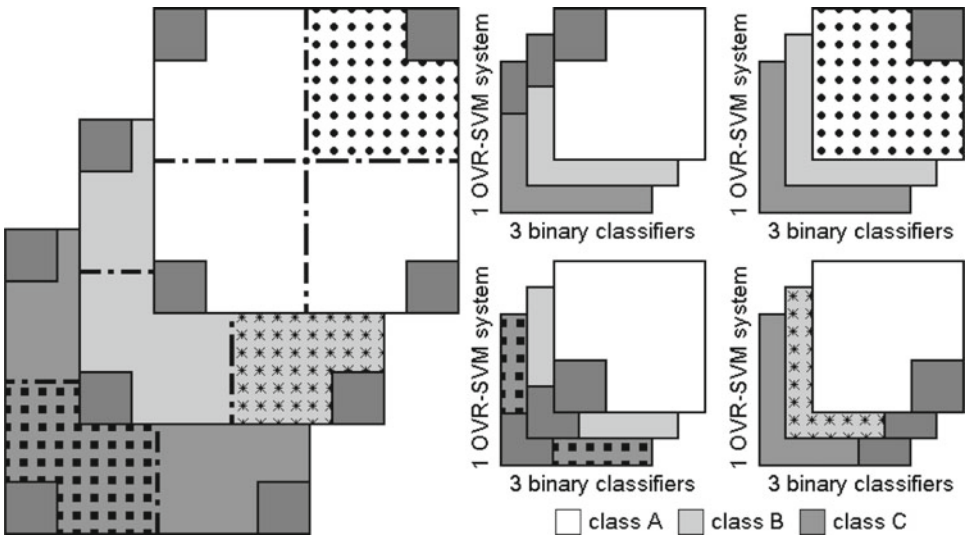

Fig. 1 Example of a 3-class classification problem and the division of the images into four regions

Eq. (2.1), the α values of the properly classified samples (True Positive, TP, and True Negative, TN) are negative. In contrast, the α values of the misclassified samples (False Positive, FP, and False Negative, FN) are positive. On the one hand, in the regions on which the α^{rt} values are positive, it has not been possible to build a good separating hyperplane between the class t and the other classes. It means that the samples of different classes are very close one to the others in the transformed feature space, and thus, it is more likely to make a mistake in the classification. On the other hand, if the α^{rt} values are negative, it means that the hyperplane built by the binary SVM classifiers accurately defines the separation between classes reducing the probability of making an error. Quality is computed as:

$$q(r,t) = \frac{\left(\sum_{i=1}^{\#\mathrm{FN}} \alpha_{\mathrm{FN},i}^{rt} + \sum_{j=1}^{\#\mathrm{FP}} \alpha_{\mathrm{FP},j}^{rt}\right) \times \mathrm{PF} + \sum_{k=1}^{\#\mathrm{TP}} \alpha_{\mathrm{TP},k}^{rt}}{\#\mathrm{FN} + \#\mathrm{FP} + \#\mathrm{TP}} \tag{3.1}$$

where #FN, #FP, #TP are, respectively, the number of false negatives, false positives and true positives, and PF is a penalty factor.

Equation (3.1) composed of two different terms:

Positive terms α values of FP and FN samples in the calibration set. Their sum is multiplied by a penalty factor (PF). The aim of the PF (PF $\in [1, \infty)$) is to penalize the classifiers that make classification errors.

Negative terms α values of TP samples.

Using Eq. (3.1), the smaller (negative) q value the better is the region to classify a class t.

TN samples have not been included in Eq. (3.1). The reason is that since the number of TN samples is usually much larger than the number of TP samples, the statistical weight of the TN samples is much higher. It can force classifiers to obtain large distances from the TN samples to the hyperplane rather than properly classify both the TN and the TP samples. For example, let's assume a 10 class problem with 10 images per class. On evaluating a classifier of the first class, in the optimum case (a perfect classifier), the number of TN is 90 (9 classes times 10 samples per class) and the number of True Positive (TP) is 10 (1 class times 10 samples). Thus, the statistical weight of the TN is 9 times the weight of the TP. Using the TN samples in Eq. (3.1), a classifier without TP (and False Negatives instead) but with high distances from the TN samples to the separation hyperplane can get a better evaluation than a classifier with 10 TP and 90 TN but with short distances. To solve this problem, the TN samples are not included in Eq. (3.1) and penalty factor (PF) is added to penalise the systems with errors.

An alternative to Eq. (3.1) can be a weighted average quality measured using the PF factor:

$$q(r,t) = \frac{\left(\sum_{i=1}^{\#\mathrm{FN}} \alpha_{\mathrm{FN},i}^{rt} + \sum_{j=1}^{\#\mathrm{FP}} \alpha_{\mathrm{FP},j}^{rt}\right) \times PF + \sum_{k=1}^{\#\mathrm{TP}} \alpha_{\mathrm{TP},k}^{rt}}{(\#\mathrm{FN} + \#\mathrm{FP}) \times \mathrm{PF} + \#\mathrm{TP}} \tag{3.2}$$

3.4 Region selection

The objective of this step is the selection of the most suitable regions for the classification of each class.

The q values of the regions computed using Eq. (3.1) are sorted and the smallest value of each class is selected as the best region to classify that class. It is possible that the same region would be the best one to classify all the different classes of the k-class classification problem or, in contrast, k different regions would be the best ones. These k regions can overlap or they can use different image pixels.

The result of the region selection is an OVR-SVM system containing the k best binary classifiers, one per each class. Figure 2 shows an example of the selection of the best binary classifiers for the image division in Fig. 1. One binary classifier has been selected from the top-right region (class A), one from the down-right region (class B) and one from the down-left region (class C). The top-left region is not suitable enough to classify any class (there are no differences between the classes) and thus, no classifiers have been selected.

3.5 Image classification

On classifying a new image, it is divided into regions and these regions are classified by their corresponding binary classifiers (in a k-class classification problem only k regions are analyzed). The image is assigned to the class of the classifier with the largest p_{value}. The credibility measure of the classification is equal to the largest p_{value} and confidence is equal to 1 minus the second largest p_{value}.

4 Classification of the Thomson scattering images

As it was mentioned previously, the TJ-II Thomson scattering CCD camera generates 2-D images from which to compute the plasma temperature and plasma density profiles. Each one of the images belongs to only one of five different categories. The class of the image depends on the experiment carried out in the TJ-II fusion device. As a consequence, the number of images available of each class is different: 42 CFF, 517 ECH, 124 BCK, 366 NBI and 223 STR for a total of 1,272 different images. Figure 3 shows an example of each one of the Thomson scattering categories.

The resolution of the images is 576×385 pixels. To speed up the region selection process, the images have been reduced

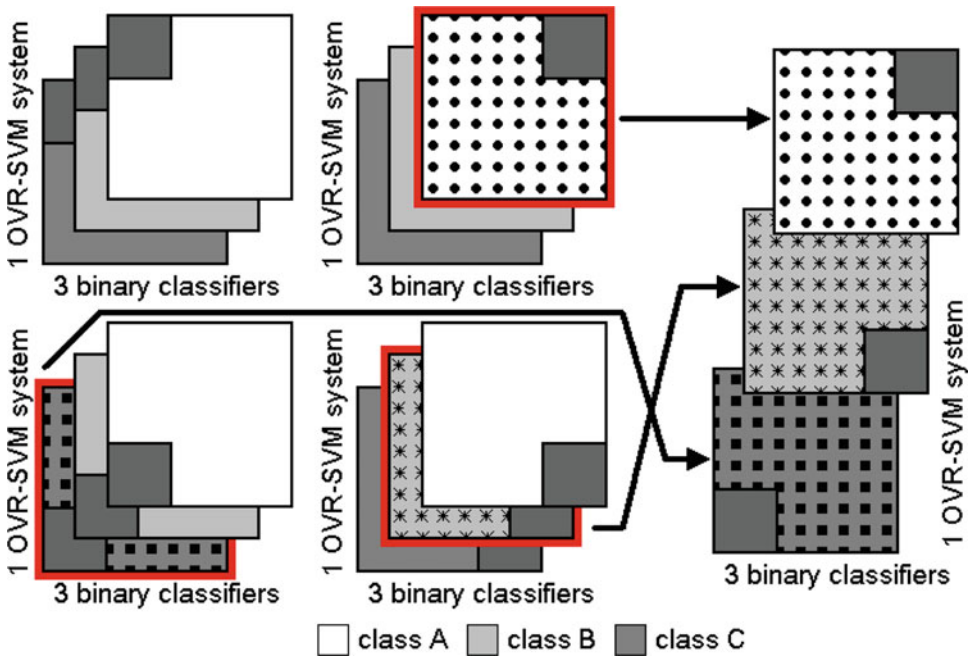

Fig. 2 Example of the region selection in a 3-class classification problem

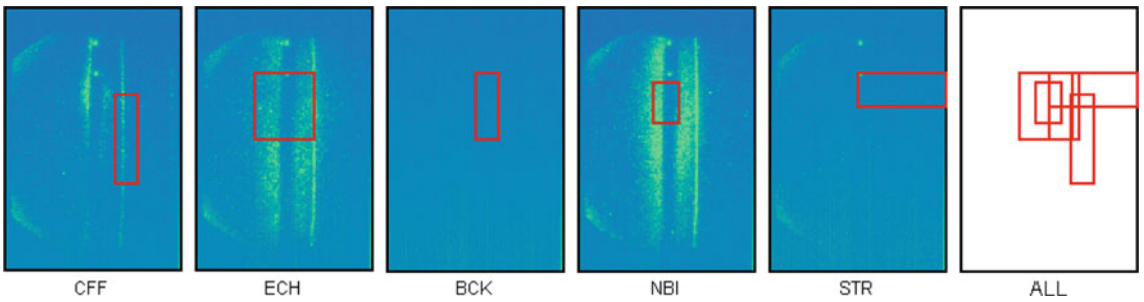

Fig. 3 Selected region of each class and the total region used by all the classes

to 144 × 97 pixels using the approximation wavelet coefficients of level 2 (Haar family) [17].

Since no expert knowledge has been used, it was not possible to divide the images using information about the problem. As an alternative, the images have been divided into rectangular regions. The height and the width of the images have been divided from 2 to 8 regions, testing all the possible combinations. A total number of 1,225 different regions have been created. Some examples of the image division can be found in Fig. 4.

The binary SVM classifiers of the OVR-SVM systems have been trained using radial-basis function (rbf) kernels [18]. The epsilon SVM parameter has been set to 1×10^{-7}. The values of the SVM complexity parameter C and the rbf kernel parameter σ are based on the problem being solved. Since the regions size and the information contained on each region are not the same, it is not possible to set and unique value of these parameters for all the regions. Therefore, five values of C ({800, 900, 1,000, 1,100, 1,200}) and five different values of σ ({8, 9, 10, 11, 12})—25 different combinations—have been tested for each region.

The total set of 1,272 images has been randomly divided into a proper training set (25 % of each class), a calibration set (25 % of each class) and a test set (50 % of each class). Therefore, the proper training set and the calibration set contain 11 CFF images, 129 ECH images, 31 BCK images, 92 NBI images and 56 STR images (319 images each one). The test set contains 20 CFF images, 259 ECH images, 62 BCK images, 182 NBI images and 111 STR images (634 images).

According to the above paragraphs, 1,225 different regions have been trained using 25 different combinations of the SVM parameters, making a total of 30,625 systems. For each class, the systems have been evaluated using Eq. (3.1), obtaining the results in Fig. 5. It is important to note that the q values of the systems have been ordered independently for each class.

Figure 5 shows the size of the selected regions as a percentage of the total image size for different penalty factor

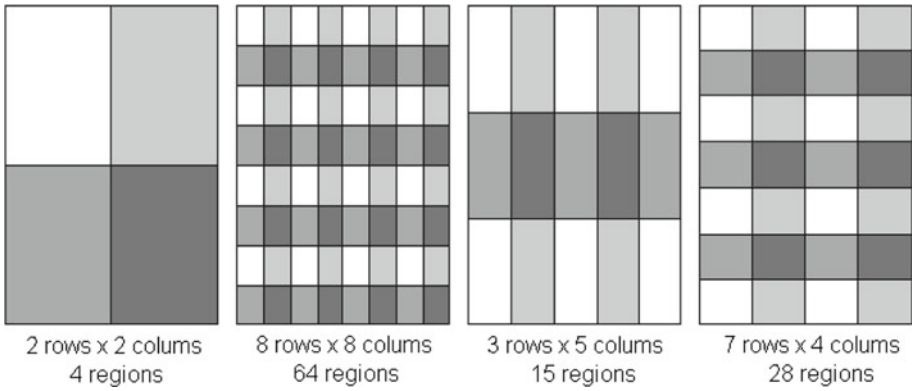

Fig. 4 Examples of the division of the Thomson scattering images in regions

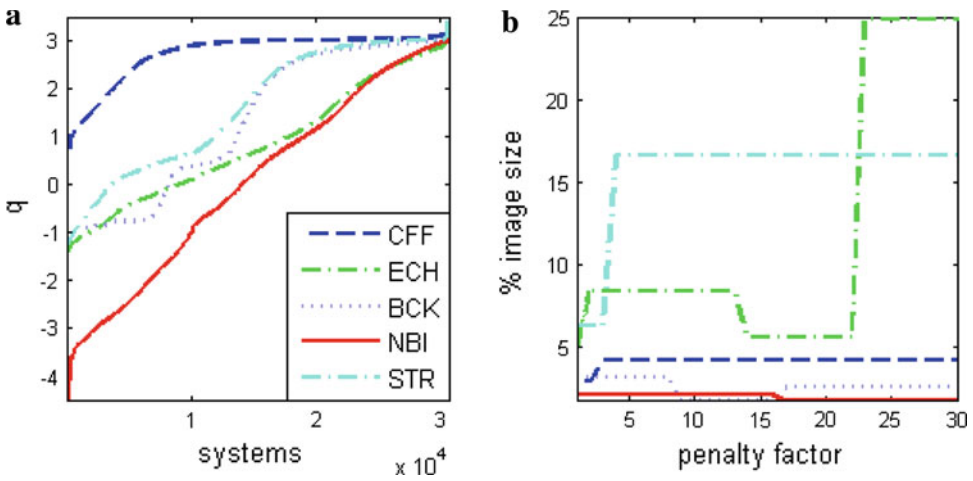

Fig. 5 Performance of the 30,625 systems trained in the experiment (**a**) and size of the selected regions for different penalty factors (**b**)

values. Large values of the penalty factor force the selected region not to make errors. It is possible to reduce the number of errors increasing the information on the selected regions and it is possible in two ways: selecting a region with more information and if it is not possible, increasing the size of the selected regions. Therefore, bigger regions are selected for larger values of the penalty factor. It is important to take into consideration that the images are divided from 2 to 8 columns/rows. Then, the maximum size of a region is 25 % (a region of an image divided into 2 columns and 2 rows) and the minimum size is 1.5625 % (8 rows and 8 columns). For large values of PF, the ECH region selected is the biggest region available.

Using PF = 3, the q value of the best systems are: CFF, −0.673; ECH, −1.0524; BCK, −1.0392; NBI, −3.1756; and STR, −1.1793. Figure 3 shows the region selected by the best systems of each class and the total area needed for all the classifiers. The total area is only a 16.49 % of the original size of the images (note that some regions are overlapped and thus the size of the total area selected is not the sum of the areas of the different classes). The success rates of the classification of the test images are CFF 15.00 %, ECH 96.53 %, BCK 100 %, NBI 98.80 %, STR 98.20 % and the total classification success rate taking into consideration the different number of test images on each class is 95.27 %. The reason of the low classification success rates of the CFF class is the small number of training examples: only 11 images.

The results obtained using the regions selected by the RSIC methodology have been compared to the ones obtained by 5 SVM one versus rest systems trained using the full images. The same proper training set, calibration set and test set described above have been used to train and test these systems. As it was mentioned previously, the full images contain more information than the regions selected by the RSIC methodology. Therefore, it is possible to obtain better classification rates than using only the RSIC regions. In this case, the total success rate using the full images is 96.37 %, slightly higher than using the RSIC regions (95.27 %). It is important to note that the RSIC regions are using only the 16.49 % of the size of the full images. Table 1 contains the detailed results obtained using both methods for each one of the image classes.

The CPU times used by the RSIC regions and the full images have been measured to quantify the reduction of

Table 1 Comparison of the classification results obtained using the RSIC regions versus the full images

Image class	Rates	Full image	RSIC
CFF 20 images	Success	11, 55.00 %	3, 15.00 %
	Errors	4, 20.00 %	5, 25.00 %
	Unknowns	5, 25.00 %	12, 60.00 %
	Region size	100 %	4.17 %
ECH 259 images	Success	246, 94.98 %	250, 96.53 %
	Errors	6, 2.32 %	6, 2.32 %
	Unknowns	7, 2.70 %	3, 1.16 %
	Region size	100 %	8.33 %
BCK 62 images	Success	62, 100 %	62, 100 %
	Errors	0, 0 %	0, 0 %
	Unknowns	0, 0 %	0, 0 %
	Region size	100 %	3.13 %
NBI 182 images	Success	181, 99.45 %	180, 98.80 %
	Errors	1, 0.55 %	0, 0 %
	Unknowns	0, 0 %	2, 1.1 %
	Region size	100 %	2.04 %
STR 111 images	Success	111, 100 %	109, 98.20 %
	Errors	0, 0 %	2, 1.80 %
	Unknowns	0, 0 %	0, 0 %
	Region size	100 %	6.25 %
TOTAL 634 images	Success	611, 96.37 %	604, 95.27 %
	Errors	11, 1.74 %	13, 2.05 %
	Unknowns	12, 1.89 %	17, 2.68 %
	Region size	100 %	16.49 %

the models complexity. 100 runs have been performed for each one of the systems. Table 2 summarizes the mean times obtained using a Intel® Core™ 2 Quad CPU Q9300 2.50 GHz, 1.95 GB RAM. Since the times of the different runs are very similar, the standard deviation values (SD) are small. The differences of the times on training the models of each class are due to the different number of images in each class. Using the RSIC regions, the proper training is 94.49 % faster than using the full images. The calibration is reduced 81.13 %. The time to complete the set is also reduced 81.83 %. The mean time to classify an image using the RSIC regions is 0.025 s versus 0.137 s using the systems trained with the full images.

Some other techniques have been previously applied to classify the Thomson scattering images. For example, [7] uses neural networks obtaining a success rate of 90.89 %. In [8], images are classified using nearest neighbours and only one image per class in the training set. It obtains a success rate of 96.87 % in a set of 165 Thomson scattering images.

Table 2 Comparison of the computation times using the RSIC regions versus the full images

	Full image		RSIC	
	Time (s)	SD (s)	Time (s)	SD (s)
Training				
CFF 11 images				
Proper training	1.206	0.026	0.075	0.001
Calibration	6.325	0.029	1.456	0.005
ECH 129 images				
Proper training	2.301	0.048	0.200	0.003
Calibration	14.398	0.071	2.024	0.004
BCK 31 images				
Proper training	1.097	0.012	0.033	0.001
Calibration	5.228	0.037	1.318	0.006
NBI 92 images				
Proper training	1.400	0.012	0.034	0.001
Calibration	8.029	0.045	1.323	0.008
STR images				
Proper training	1.712	0.012	0.124	0.001
Calibration	10.257	0.056	1.600	0.074
TEST—634 images	86.937	0.696	15.802	0.6958

5 Discussion

A new methodology for region selection and image classification has been introduced. Using this methodology, it is possible to locate the most significant regions in images for a classification problem (the regions that allow a predictor to classify a new image, i.e. the regions with unique information of each class) and to classify new image using only a small percentage of the original images size obtaining high success classification rates. The methodology is based on a novel interpretation of the non-conformity measure that allows identification of the best region to distinguish a class versus the rest of classes of a multi-class classification problem. It has been applied to the identification of the most important regions of the images generated by a diagnostic of a fusion device. High classification success rates, a significant reduction of the original image sizes and a reduction in the training and test CPU times have been obtained.

In this approach, only one region per class is selected as the most suitable one to classify that class. It is possible that, for very complex classification problems, several regions would be needed. On classifying a new image using several regions per class, some different approaches can be used, for example:

- The new image is assign to the class of the region with the largest p_{value}. This option is similar to the one in this paper but using more regions and thus more p_{values}.

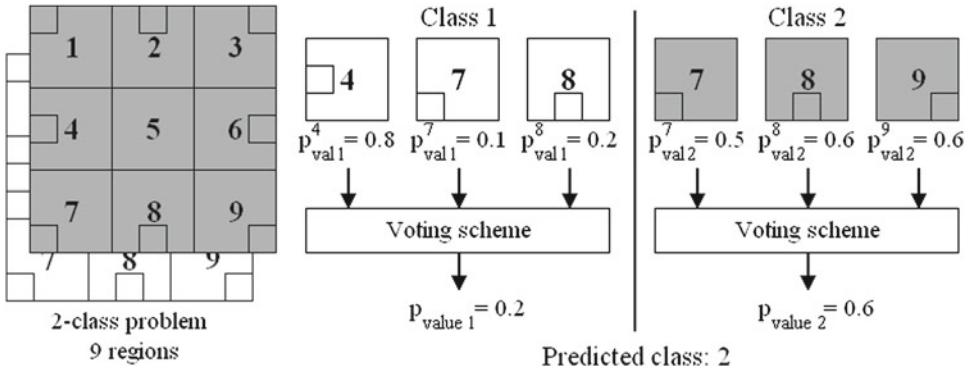

Fig. 6 Example of a classification with multiple regions and a voting schema

- Implementing a voting scheme for the regions of the same class and then use the output of the voting schemes as the p_{values} of each class. Figure 6 shows an example of this alternative using a 2-class classification problem and 9 regions. Three regions have been selected to classify each class. In order to select a p_{value} for each class, a voting scheme is used. Then, the resulting p_{values} are compared as described in Sect. 3.5. In this case, the predicted class is 2, the credibility is 0.6 and the confidence is 0.8. It is important to note that although the largest p_{value} is the one of the region 4 of class 1, since the p_{values} of the regions 7 and 8 are low, the voting scheme has obtained a low p_{value}. If there would not be a voting scheme, the predicted class would be 1.

References

1. Gammerman, A., Vovk, V.: Hedging predictions in machine learning. Comput. J. **50**(2), 151–163 (2007). doi:10.1093/comjnl/bxl065
2. Barth, C.J., Pijper, F.J., Meiden, H.J.v.d., Herranz, J., Pastor, I.: "High-resolution multiposition Thomson Scattering for the TJ-II stellarator". Rev. Sci. Instrum. **70**(1), 763–767 (1999)
3. Herranz, J., Castejón, F., Pastor, I., McCarthy, K.J.: The spectrometer of the high-resolution multiposition Thomson scattering diagnostic for TJ-II. Fusion Eng. Design. **64**(4), 525–536 (2003)
4. Sanchez, J., et al.: Overview of TJ-II experiments. Nuclear Fusion **51**, 094022 (2011) (10 pp). doi:10.1088/0029-5515/51/9/094022
5. Vega, J., et al.: Application of intelligent classification techniques to the TJ-II Thomson Scattering diagnostic. In: Proceedings of the 32nd EPS conference on plasma Phys. ECA vol. 29C, P-2.090. Tarragona, 27 June-1 July (2005)
6. Makili, L., Vega, J., Dormido-Canto, S., Pastor, I., Pereira, A., Farias, G., Portas, A., Perez-Risco, D., Rodriguez-Fernandez, M.C., Busch, P.: Upgrade of the automatic analysis system in the TJ-II Thomson Scattering diagnostic: New image recognition classifier and fault condition detection. Fusion Engineering and Design, vol. 85, issues 3–4, Proceedings of the 7th IAEA Technical Meeting on Control, Data Acquisition, and Remote Participation for Fusion Research, pp. 415–418, July 2010, ISSN: 0920-3796. doi:10.1016/j.fusengdes.2009.10.004
7. Farias, G., Dormido, R., Santos, M., Duro, N.: Image classifier for the TJ-II Thomson scattering diagnostic: evaluation with a feed forward neural network. Lect. Notes Comput. Sci. **3562**, 362–381 (2005). doi:10.1007/11499305_62
8. Vega, J., Murari, A., Pereira, A., Gonzalez, S., Pastor, I.: Accurate and reliable image classification by using conformal predictors in the TJ-II Thomson scattering. Rev. Sci. Instrum. **81**, 10E118 (2010). doi:10.1063/1.3478689
9. Makili, L., Vega, J., Dormido-Canto, S., Pastor, I., Murari, A.: Computationally efficient SVM multi-class image recognition with confidence measures. Fusion Eng. Design, corrected proof, Available online 22 March 2011, ISSN: 0920-3796 (in press). doi:10.1016/j.fusengdes.2011.02.081
10. Daubechies, I.: Ten lectures on wavelets. CBMS-NSF conference series in applied mathematics. SIAM Ed (1992)
11. Saunders, C., Gammerman, A., Vovk, V.: Transduction with confidence and credibility. Proc. IJCAI'99 **2**, 722–726 (1999)
12. Vapnik, V.: Statistical learning theory. Wiley, INC.m, London (1998)
13. Papadopoulos, H.: Inductive Conformal Prediction: Theory and Application to Neural Networks. Tools in Artificial Intelligence, pp. 315–330. I-Tech, Vienna (2008)
14. Shafer, G., Vovk, V.: A tutorial on conformal prediction. J. Mach. Learn. Res. **9**, 371–421 (2008)
15. Papadopoulos, H., Proedrou, K., Vovk, V., Gammerman, A.: Inductive confidence machines for regression. In: Proceedings of the 13th European Conference on Machine Learning (ECML'02). Lecture Notes in Computer Science, vol. 2430, pp. 345–356. Springer, Berlin (2002)
16. Weston, J., Watkins, C.: Support vector machines for multi-class pattern recognition. ESANN Proc. 219-224 (1999), ISBN: 2-600049-9-X
17. Mallat, S.: A wavelet tour of signal processing, 2nd edn. Academic Press, New York (2001)
18. Cherkassky, V., Mulier, F.: Learning from data: concepts, theory and methods, 2nd edn. Wiley, New York (2007)