APPLICATION OF DATA FUSION THEORY AND SUPPORT VECTOR MACHINE TO X-RAY CASTINGS INSPECTION

Ahmad Osman[1*], Valérie Kaftandjian[2], Ulf Hassler[1]

[1] Fraunhofer Development Center X-ray Technologies, A Cooperative Department of IZFP Saarbruecken and IIS Erlangen, Dr-Mack-Str. 81, 90762 Fuerth, Germany.
[2] National Institute of Applied Sciences INSA- Lyon, Non Destructive Testing using Ionising Radiations laboratory CNDRI, Bat. St. Exupery, 20 Avenue Albert EINSTEIN, 69621 Villeurbanne, France.

1. Introduction:

X-ray inspection is a traditional non-destructive testing method used to thoroughly test industrial parts, such as aluminum castings in the automotive sector. Safety specifications and quality control task are the main focus of the inspection process. Digital image processing, computational intelligence and hardware progress allowed automating this task. While the detection of true defects is the objective, one main difficulty in X-ray inspection is the detection of false alarms (or false defects), especially if very small and low contrasted defects have to be detected. Therefore, reducing the rejection ratio of good parts without risking missing true defects is a serious challenge.

The automatic detection and recognition of defects requires computerized image processing, image analysis and decision process. The image processing step is critical to detect potential defects. During the image analysis step features are extracted to be further used to differentiate between true defects TD and false defects FD. This is where our paper intervenes in order to discriminate TD from FD. A specific approach based on data fusion is developed to combine different features with each other, each feature being considered a source of information. Furthermore, Support Vector Machine is used as another classification method and a comparison between the two techniques is presented.

2. Data fusion classification method

Data fusion is widely used in several fields including medical diagnostics, human recognition and industrial applications to combine information provided by different sources. The objective is to identify and classify an element x into a class C_i using f_j information provided by different sources.
The most used models in data fusion are the probability approach, the possibility approach and the evidence theory for data fusion. The choice of one of them is dependent upon the ease of its implementation and especially its ability to model uncertain and (or) imprecise knowledge. Common difficulties exist for all these approaches:

- Definition of the frame of discernment.
- Modelling of the knowledge (expert and statistic).
- Definition of the combination rules.
- Choice of the decision criteria.

Dempster-Shafer evidence theory (DS) was developed as an attempt to generalize probability theory (Dempster 67, Shafer 76). It is suitable to reason with uncertainty and allows us to

* Corresponding author: Tel. +49 911 58061 7642. Email: ahmad.osman@iis.fraunhofer.de (A. Osman).

distinguish between uncertainty and imprecision. This is achieved in particular by making it possible to handle composite hypotheses. DS theory is also suitable for combining information from different sources. In DS theory, the frame of discernment Θ is formed of N subsets A_i which can be a simple hypothesis H_i or a union of simple hypotheses. These hypotheses are not necessarily exclusive, e.g. {friend, enemy, neutral}. Θ represents the working space for the application being considered since it consists of all propositions for which the information sources can provide evidence. Information sources can distribute mass values on subsets of the frame of discernment. Obtaining the mass distribution or function $m(A_i)$ ($0 \le m(A_i) \le 1$) is the most important step since it represents the knowledge about the current application as well as the uncertainty and imprecision incorporated in the selected information source.

In our work the frame of discernment is formed of 3 hypotheses: hypothesis H_1 ("this object is a true defect TD"), hypothesis H_2 ("this object is not a true defect", i.e. it is a false defect FD), and the ignorance is represented by the combined hypothesis ($H_1 \cup H_2 = H_3$). Further in this paper, class A stands for the hypothesis H_1 and class B stands for the hypothesis H_2.

In previous works, the data fusion approach was used with the aim to improve the detection of weld defects in (Kaftandjian et al, 2002), and in castings inspection (Lecomte et al., 2006). The obtained results proved to be precise and more reliable decisions were obtained. However, supervision by an expert was necessary to assign the confidence levels (or mass values).

In the present work, the mass value attribution is completely automated, and the expert supervision is no longer necessary. The method, introduced in (Osman et al., 2009), allow converting from the space of feature values into the mass values space. This method is divided into two processes: learning process and validation process**.**

2. 1 Learning process:

The spatial repartition of feature's values is divided into regions of confidence. To elaborate these regions, the global histogram of class A (TD) and class B (FD) is used.

Firstly, this histogram is divided into a set I of intervals. For each interval $i \in$ I, the percentage of TD (instances of the class A) present in this region is calculated using the following function:

$$P_{A,B}(i) = \frac{h_A(i)}{h_A(i) + h_B(i)}$$

$h_A(i)$ represents the number of instances of A inside *i.*

$h_B(i)$ represents the number of instances of B inside *i.*

This percentage is assigned as mass to H_1: $m(H_1) = P_{A,B}(i)$.

Null is the mass assigned to the hypothesis H_2: $m(H_2) = 0$.

1-$m(H_1)$ is the mass assigned to H_3: $m(H_3) = 1 - P_{A,B}(i)$.

Secondly, subsequent intervals are congregated to form a region of confidence. In this step, the first constraint on the variation of $P_{A,B}(i)$ between two adjacent regions is used: a fixed threshold DV [first constraint] is applied to $\Delta P_{A,B}(i)$.

If $\Delta P_{A,B} = \left| P_{A,B}(i+1) - P_{A,B}(i) \right| < DV$: Region *i* is merged with region *i+1*, they will have in this case the same mass values.

The influence of this threshold *DV* (called Derivation Variation) on the system performance and stability is studied.
DV changes between 0 and 1. The number of regions of confidence decreases with the increase of *DV* since the merging of intervals will increase with a high *DV.*
At the end of this step, some obtained regions contain a very small number of points to be considered as significant. Therefore a second constraint on the number of points existing in each region is imposed: a region should contain at least a certain percentage of points *Perc*, which we specify, [second constraint] to be considered as having enough significance. Let M be the region which contain the biggest number of points N_M inside, the minimal number of points to be respected inside each region is:
$N_c = Perc \cdot N_M$
This influence of *Perc* on the system's performance is also studied.

The theory of fuzzy sets is used to ensure a continuous transition between the regions of confidence by introducing membership functions.

After the estimation of regions of confidence and their corresponding mass values and membership functions, the fusion process for combining different features takes place. We first calculate the single mass values for the objects of the learning database. After that we combine the mass values given by the features (two features fusion, three best features fusion, all features fusion) and also we will use the statistical data fusion: mean mass, median mass.

To classify an object using the information source f_k, a threshold S is applied on its mass value $m(H_1)$. The object is classified as:
- a defect if $m(H_1) \geq S$
- unknown (defect or not) if $m(H_1) < S$

The influence of this threshold S is also studied.
The classification results are then compared to the true decision given by the expert, and the following rates are computed:

- Percentage of correct decisions PCD:

$$PCD = \frac{\text{number of true defects correctly classified + false defects correctly classified}}{\text{total number of true defects and false defects}}$$

- True Defects detection ratio PTD:

$$PTD = \frac{\text{number of true defects correctly classified}}{\text{total number of true defects}}$$

- False Defects detection ratio PFD:

$$PFD = \frac{\text{number of false defects correctly classified}}{\text{total number of false defects}}$$

- Overall detection ratio R:

$$R = \frac{a \cdot PCD + b \cdot PTD + c \cdot PFD}{a + b + c}$$

The use of a, b and c to compute the overall detection rate R is driven by industrial requirements. It is very important in the castings industry to detect as many real defects as

possible, while preventing the false alarm rate. Thus, the overall rate R is computed with a = c = 1, and b = 5

Two possibilities are present to select the best fusions of features (best combinations):

1. The best combinations are chosen relatively to the original external inspection system, in our case i.e. ISAR decision, where ISAR (Intelligent System for Automated Radioscopy) is developed by Fraunhofer EZRT, and is used for radioscopic quality control in the production of castings. Some information about the ISAR system can be found in the reference (Fuchs, 2003).

2. The user indicates the required PCD, PTD and PFD to consider a source as member of the best combinations. For this source PCD, PTD and PFD should be equal or higher then the asked for performance.

The first option was already used in previous work (Osman et al., 2009). In this paper we will use the second option.

2.2 Optimization of our method

Principally the influence of three parameters needs to be studied: the derivation variation *DV*, the minimal number of points inside a confidence region *Nc* (depends on *Perc*) and the threshold S applied to the mass value of an object to be classified as a defect. By default these values were set to 0.2 for *DV*, 0.1 for *Perc* and S varies between 0.6 and 0.9.

The response of the system to the variation of *DV*, *Nc* or S is evaluated by measuring the number of successful sources NSS and the performance of optimal overall sources obtained from a certain regulation. A source is successful if the ratio of detection of class A is PTD> = 0.9, the ratio of detection of class B is PFD $\geq$ 0.8 and the true decision ratio is PCD $\geq$ 0.85.

This study is performed on a learning database formed of 115 TD (instances of class A) and 65 FD (instances of class B). The results are validated on a testing database formed of 116 TD and 65 FD. Eleven features (Area, Depth, InOutContrast…) are extracted from each potential defect, further to be used in the classification process. On the learning database ISAR possess an R = 0.925, PTD = 0.974 and PFD = 0.723. On the testing database ISAR gives the following results: R = 0.932, PTD = 0.982 and PFD = 0.723.

2.2.1 Variation of *DV* and S:

While *DV* changes *Perc* is fixed to 0.1. *DV* has direct influence on the number of the regions of confidence. In term of number of successful sources NSS, the result is shown in figure 1.

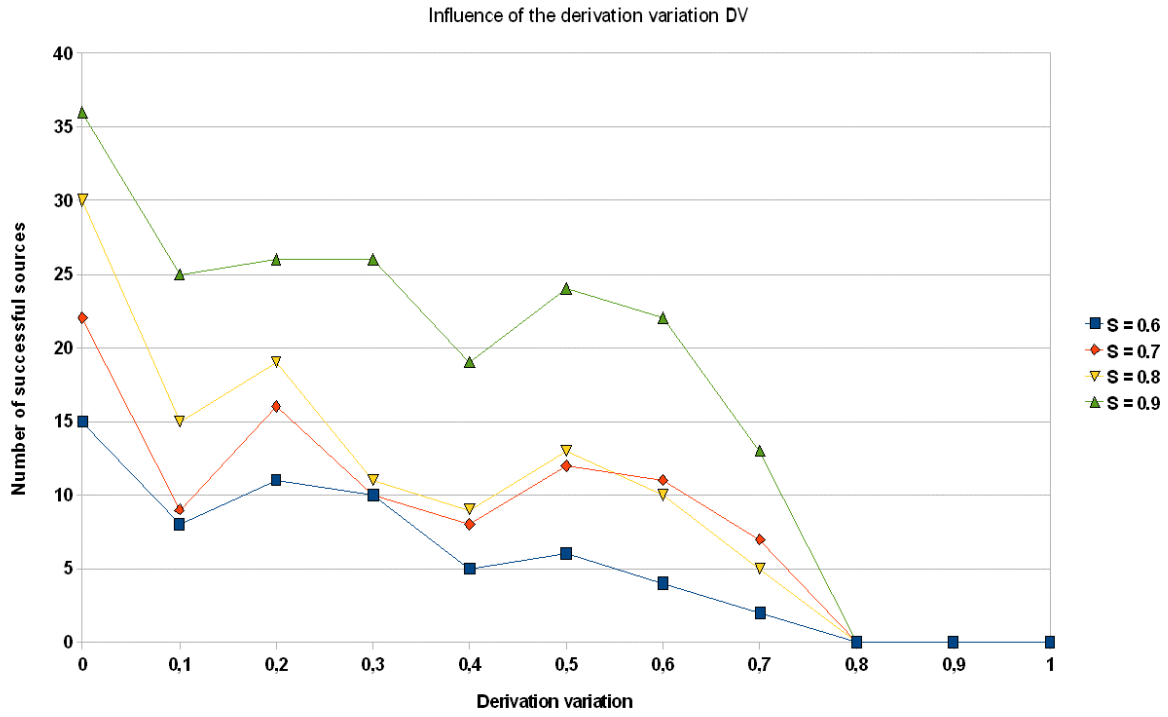


Figure 1. Influence of the derivation variation and S on the NSS.

Generally the number of successful sources NSS is higher when *DV* is small. It also increases when S gets higher. This fact is due to the optimistic Dempster-Shafer fusion rule. For a *DV* higher than 0.7, NSS is null, therefore no optimal sources are available.
The optimal overall R variation is presented in figure 2. The best optimal overall R achieved is R = 0.992 with PTD = 0.991 and PFD = 1 for S = 0.6. This source is the Mean mass.

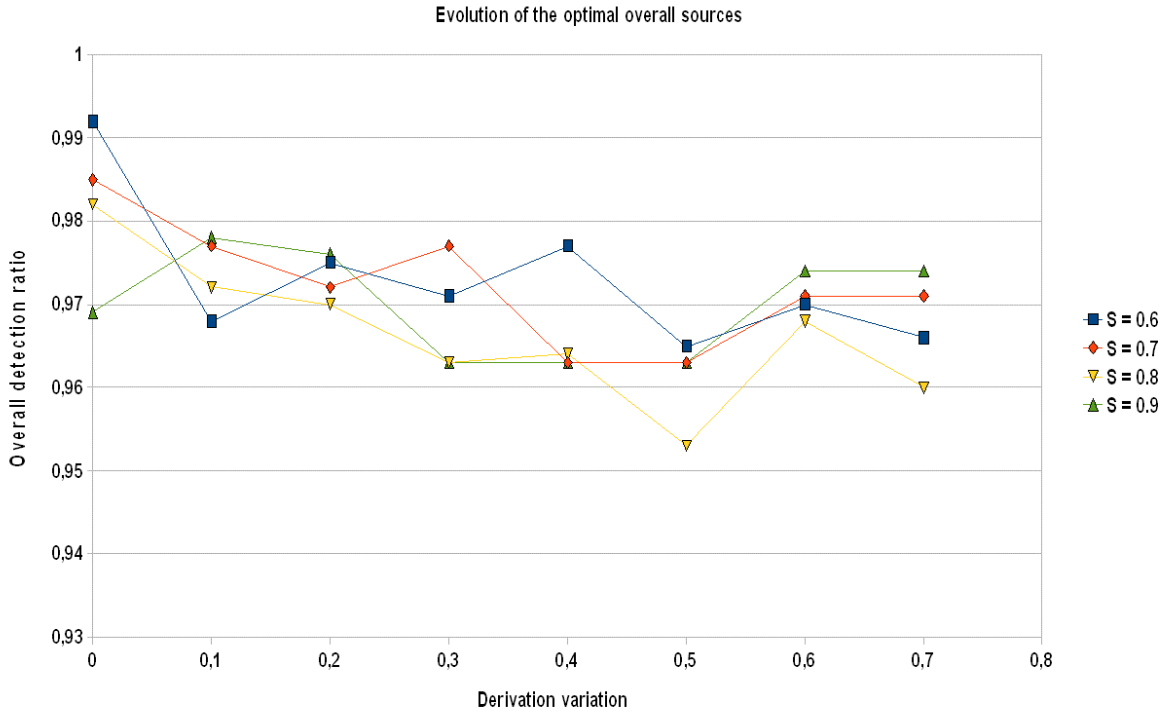


Figure 2. Influence of the derivation variation and S on the overall detection ratio.

2.2.2 Variation of *Perc* and S:

DV is fixed to 0.2 while *Perc* changes. *Perc* is changing between 0 and 1 and S is changing between 0.6 and 0.9.
NSS decreases when *Perc* increase. For the same *Perc*, a higher threshold on mass values S gives more successful sources (see figure 3).

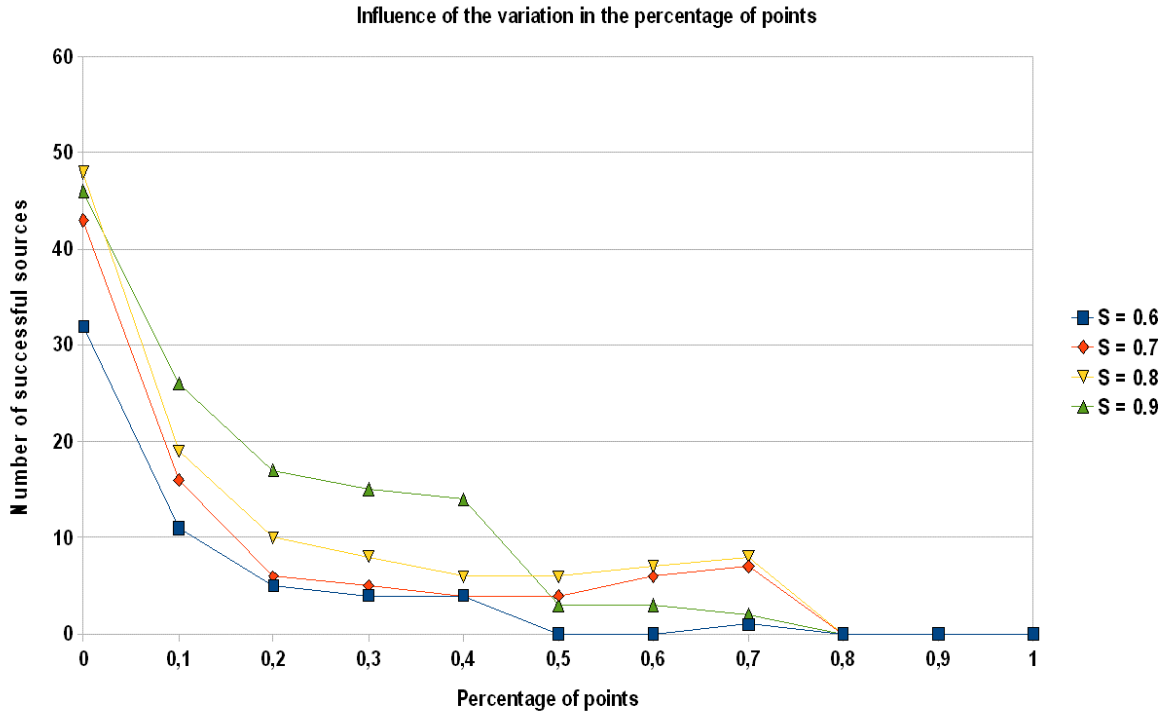


Figure 3. Influence of the variation of *Perc* and S on the NSS.

As for the optimal overall sources, see figure 4, it is obvious that the lower the constraint on the percentage of points inside a region is the better the detection ratio is. This is due to the higher number of regions of confidence that are found when *Perc* is small.
The best optimal overall R achieved is R = 0.988 with PTD = 1 and PFD = 0.938. Two sources give this result:
Dempster-Shafer fusion of MaxElongation and InOutContrast for S = 0.6.
Dempster-Shafer fusion of Depth2Thickness and InOutContrast for S = 0.9.

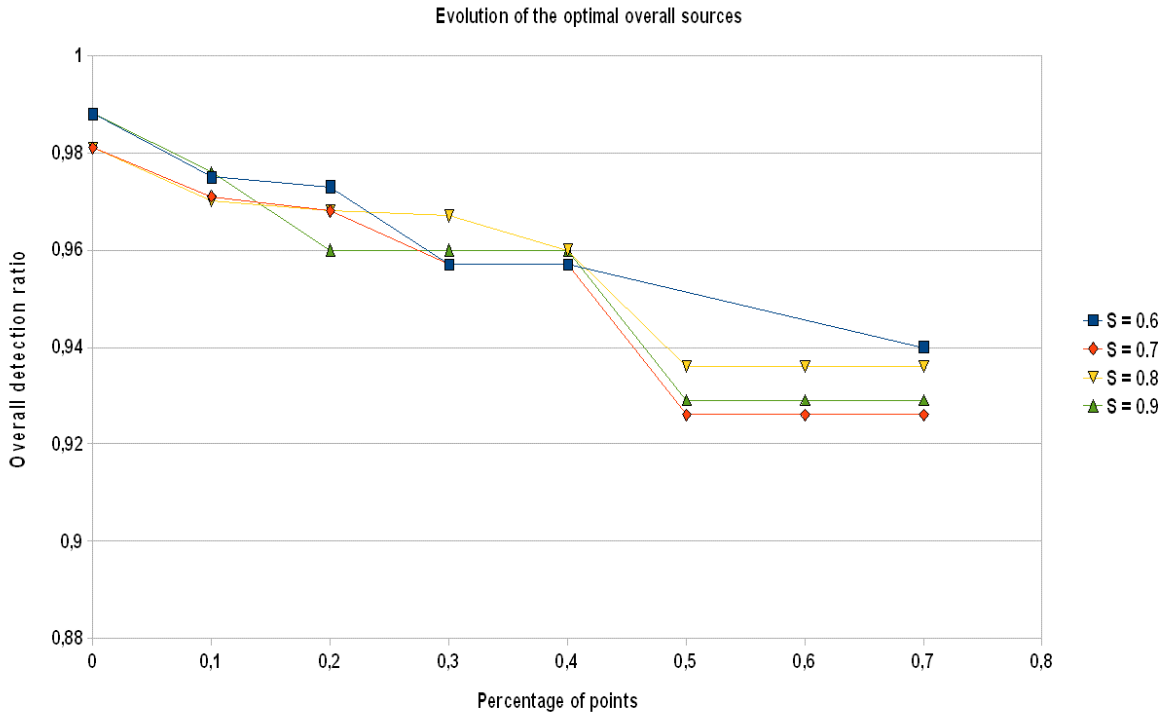


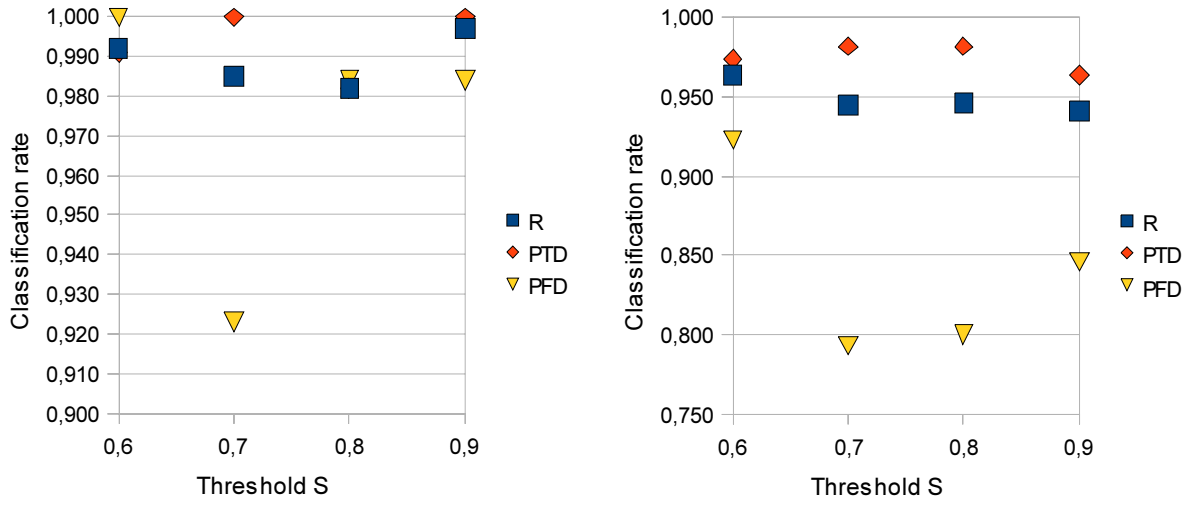
Figure 4. Influence of the variation of *Perc* and S on the overall detection ratio.

2.2.3 Variation of S for optimal *DV* and *Perc*:

Considering the results obtained the best values of *DV* and *Perc* seem to be 0 on the learning database. Thus, for *DV* = 0 and *Perc* = 0, the threshold is modified from 0.6 to 0.9. The number of successful sources remains high and varies between 33 and 51. The evolution of the classification rates is shown in figure 5.a. For each threshold S, the optimal overall source is different (see table 1). This last source gives the best result with a ratio R = 0.997.

Source	Parameters	R	PTD	PFD
Mean Mass	*DV* = 0, *Perc* = 0, S = 0.6	0.992	0.991	1
DS fusion InOutContrast & InOutContrastGV	*DV* = 0, *Perc* = 0, S = 0.7	0.985	1	0.923
DS fusion InOutContrast & MaxDepth	*DV* = 0, *Perc* = 0, S = 0.8	0.982	0.982	0.984
DS fusion Depth2Thcikness & InOutContrast	*DV* = 0, *Perc* = 0, S = 0.9	0.997	1	0.984

Table 1. Evolution of the classification rates as a function of threshold S, for *DV* = 0 and *Perc* = 0.



(5.a) (5.b)

Figure 5. Evolution of the classification rates as a function of threshold S, for $DV = 0$ and Perc = 0 (5.a) on the learning database (5.a) and corresponding performances on the testing database (5.b). (NB: only the first four sources of table 2 are used in (5.b)).

2.3 Validation process

The validation process is conducted on the testing database. Table 2 presents the results achieved with corresponding settings of the parameters *DV*, *Perc* and S.

Source	Parameters	R	PTD	PFD
Mean Mass	$DV = 0, Perc = 0, S = 0.6$	0.964	0.974	0.923
DS fusion InOutContrast & InOutContrastGV	$DV = 0, Perc = 0, S = 0.7$	0.945	0.982	0.793
DS fusion InOutContrast & MaxDepth	$DV = 0, Perc = 0, S = 0.8$	0.946	0.982	0.8
DS fusion Depth2Thickness & InOutContrast	$DV = 0, Perc = 0, S = 0.9$	0.941	0.964	0.846
Mean Mass	$DV = 0, Perc = 0.1, S = 0.6$	0.97	0.974	0.953
DS fusion MaxElongation & InOutContrast	$DV = 0.2, Perc = 0, S = 0.6$	0.955	0.982	0.841
DS fusion Depth2Thickness & InOutContrast	$DV = 0.2, Perc = 0, S = 0.9$	0.949	0.982	0.815

Table 2. Classification rates on the testing database using the best overall sources issued from the evaluation of *DV*, *Perc* and S on the learning database.

When comparing figure 5.a and 5.b, one can see that the PFD rate gets lower in the testing database, and is mainly responsible for the decrease of the overall classification R ratio.

3. Comparison with SVM

Support Vector Machine (SVM) is a popular technique for numerical data classification. This type of learning algorithm, introduced in the 1990s, is based on results from statistical learning theory (Schölkopf, 2002). The basic concept of this technique is to find an essential subset of the gained sample data. Support Vector Machines can be used to classify elements in a certain feature space. They work in a two step process. The first is the training (with representative sample data) where the support vectors are generated. The second step is the regression/classification of unknown data in the feature space. Support Vector Machines can handle two or more classes. A detailed theoretical introduction to SVM can be found in (Niemann, 1983) and a good overview of two categories classification using SVM is presented in (Burges, 1998).
The same learning database used above is used to train the SVM. The testing process is also conducted on the same testing database.

The performance of the SVM on the learning database is:
PTD = 0.982, PFD = 1 and R = 0.985.

The performance of the SVM on the testing database is:
PTD = 0.965, PFD = 0.969697 and R = 0.966331.

4. Discussion and conclusion

The following table and figure synthesize the results in terms of R ratio.

Source	Parameters	Learning	Testing
1. ISAR		0,925	0,932
2. SVM		0.985	0.966
3. Mean Mass	$DV = 0$, $Perc = 0.1$, S = 0.6	0.992	0.970
4. Mean Mass	$DV = 0$, $Perc = 0$, S = 0.6	0.992	0.964
5. DS fusion Depth2Thickness & InOutContrast	$DV = 0$, $Perc = 0$, S = 0.9	0.997	0.941
6. DS fusion MaxElongation & InOutContrast	$DV = 0.2$, $Perc = 0$, S = 0.6	0.988	0.955
7. DS fusion Depth2Thickness & Volume	$DV = 0.2$, $Perc = 0.1$, S = 0.8	0.97	0.9559

Table 3. Achieved performances using SVM, ISAR and data fusion classifiers.

The best source selected from the learning stage is not the best after testing. This is particularly important for the source 5 which is the best source obtained for the parameters ($DV = 0$ and $Perc = 0$). For those parameters, the regions of confidence are as close as possible to the learning data (because $DV = 0$ implies that all intervals of the histogram are kept as regions of confidence). This induces the best result at the learning stage (R = 0.997), but on the other hand the regions of confidence are less adapted to the testing database. Finally a modelling of the features histogram with less regions (such as source 7 obtained with $DV = 0.2$) yields a better result on the testing database for Dempster-Shafer fusion (R = 0.959).

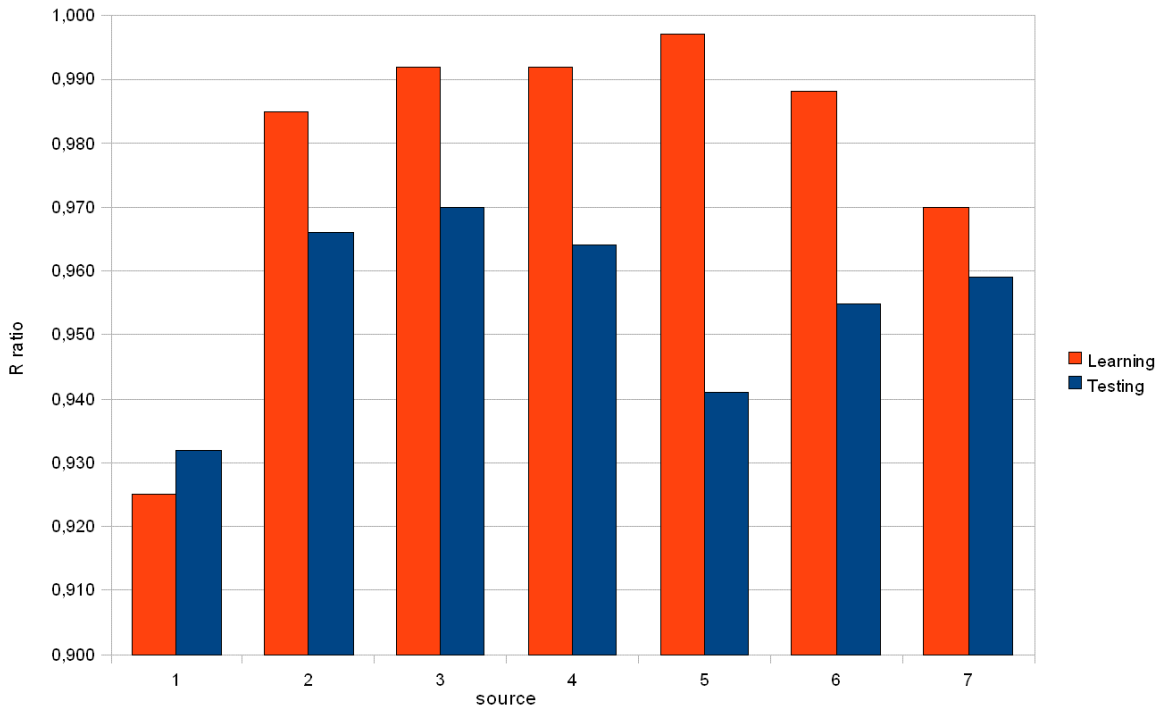


Figure 6: Performance (R ratio) of different classifiers on learning and testing.

References

Burges C., "A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery", Vol. 2, No. 2, pp.121-167, 1998.

Dempster A., "Upper and lower probabilities induced by multivalued mapping", Annals of Mathematical Statistics, 38, pp. 325-339, 1967.

Fuchs T., Hassler U., Huetten U., and Wenzel T., "A new system for fully automatic inspection of digital flat-panel detector radiographs of aluminium castings". Proceedings of 9th European Conference on Non-Destructive Testing (ECNDT), Sept. 25 – 29, 2006, Berlin Germany.

Kaftandjian V., Dupuis O., Babot, D. and Zhu Y., "Uncertainty modeling using Dempster-Shafer theory for improving detection of weld defects". Pattern Recognition Letters, Volume 24, pp. 547-564, 2003.

Osman A., Kaftandjian V., Hassler U., "Improvement of X-ray castings inspection reliability by using Dempster-Shafer data fusion theory". Submitted to Pattern Recognition Letters, 2009.

Niemann. H., "Klassifikation von Mustern", Springer Verlag, 1983.

Shafer G., "A mathematical theory of evidence", Princeton University Press, Princeton, pp. 297, 1976.

Schölkopf B. an Smola A.J., "Learning with kernels", MIT Press, 2002.