

Preprint. Please do not cite this document, use instead:

Pereda, M., Santos, J.I., Martín, Ó., Galán, J.M. (2015) Direct quality prediction in resistance spot welding process: Sensitivity, specificity and predictive accuracy comparative analysis. *Science and Technology of Welding and Joining*, Volume 20, Issue 8 (November 2015), pp. 679-685.

DOI: <http://dx.doi.org/10.1179/1362171815Y.0000000052>

Direct quality prediction in resistance spot welding process: Sensitivity, specificity and predictive accuracy comparative analysis

María Pereda^{a,*}, José Ignacio Santos^a, Óscar Martín^b, José Manuel Galán^a

^a INSISOC, Área de Organización de Empresas, Departamento de Ingeniería Civil, Escuela Politécnica Superior, Universidad de Burgos, Edificio La Milanera, C/Villadiego S/N, Burgos 09001, Spain.

^b Ingeniería de los Procesos de Fabricación, Departamento CMeIM/EGI/ICGF/IM/IPF, Universidad de Valladolid, Escuela de Ingenierías Industriales, Paseo del Cauce 59, Valladolid 47011, Spain.

* Corresponding author. Ph.: +34-947259431, Fax: +34- 947259478, E-mail: mpereda@ubu.es (M. Pereda)

Abstract In this work, several of the most popular and state-of-the-art classification methods are compared as pattern recognition tools for classification of resistance spot welding joints. Instead of using the result of a non-destructive testing technique as input variables, classifiers are trained directly with the relevant welding parameters, i.e. welding current, welding time and the type of electrode (electrode material and treatment). The algorithms are compared in terms of accuracy and area under the receiver operating characteristic (ROC) curve metrics, using nested cross-validation. Results show that although there is not a dominant classifier for every specificity/sensitivity requirement, support vector machines using radial kernel, boosting and random forest techniques obtain the best performance overall.

Keywords: Resistance Spot Welding; Classification; Pattern recognition; Quality Control; Support Vector Machines; Random Forest; Artificial Neural Networks.

1 Introduction

Resistance spot welding (RSW) is a simple and cost-effective manufacturing process¹ which is extensively used for joining sheet steel in the automobile industry due to its high speed and adaptability for automation.²

The number of RSW joints per vehicle is very high (usually varying between 3000 and 7000)³ and the tendency in the highly competitive automobile industry is to reduce it as much as possible. Thus, the development of decision support tools that can assist in a fast, flexible and efficient procedure to classify RSW joints according to their quality level can be of great interest.^{4,5} More specifically, the aim of the present work is to identify, among several options, the method that gives the best results in predicting the quality level of RSW joints directly from welding parameters,^{6,7} such as welding time, welding current, electrode material and treatment applied to electrode material.⁸ Such a prediction tool could be used to identify optimal values of the welding parameters and, consequently, to reduce the workload and the cost of post-production quality controls significantly.

In this work classifiers are compared from different perspectives. Initially, a single measure of accuracy (i.e. the classification rate) is calculated for every classifier. However, there are two types of errors that a predictive tool can produce: false positives (type I errors) and false negatives (type II errors). Thus, as pointed out by Bradley,⁹ misclassification rate –which confounds both types of errors– is not necessarily the objective function to minimize, but rather misclassification cost, which gives a different weight to the different types of errors. In the automobile industry, a satisfactory RSW joint wrongly classified as deficient will make monetary costs increase, but a deficient joint wrongly classified as acceptable will compromise safety. Consequently, classifiers in this paper have been analysed taking both types of errors into account, by using the Receiver Operating Characteristic (ROC) curve together with the Area Under the Curve (AUC) measure. This approach gives a general perspective of the performance of the classifier for the complete range of operational points or decision thresholds taking into account the trade-off between the two types of errors. Additional comparisons are also offered for fixed values of specificity and sensitivity. The analysis is completed using statistical tests (ANOVA, Duncan's new multiple range test and Waller-Duncan's) to compute the significance of the results.

2 Experimental methods

2.1 Materials and equipment

The material welded by the RSW process is sheet steel, whose chemical composition is shown in Table 1. The thickness of the sheet steel is 1 mm.

Table 1. Chemical composition of the sheet steel (wt. %).

C	Mn	Si	P	S	Al
0.05	0.26	0.02	0.012	0.011	0.033

The steel sheets are welded in a single-phase alternating current (AC) 50 Hz equipment by using water-cooled truncated cone electrodes with 5 mm face diameter.

2.2 Welding conditions

A total of 330 joints are obtained by the RSW process. Among the different RSW parameters,^{5,10} welding time, welding current, electrode force, electrode material and treatment applied to electrode material were considered. The values recommend by McCauley et al.¹¹ are taken as reference for the first three parameters (see Table 2); electrode force is kept constant for all RSW joints; whilst welding time and welding current take different values avoiding disturbances such as expulsion^{5,12}. The different electrode materials and treatments are also shown in Table 2.

Table 2. RSW parameters.

Welding time (s)	Welding current (kA RMS)	Electrode force (N)	Electrode material (RWMA Group A ^(a))	Treatment ^(d) (Applied to electrode material)
0.08				
0.10				
0.12	4			
0.14	5		Class 2 ^(b)	O61 ^(e)
0.16	6	980.7	Class 3 ^(c)	TH02 ^(f)
0.20	7			TF00 ^(g)
0.24	8			
0.28				
0.32				

0.36

0.4

- (a) Copper-base alloy.¹¹
 - (b) Chromium-copper alloy.¹¹
 - (c) Beryllium-cobalt-copper alloy.¹¹
 - (d) Temper designations according to ASTM B 601-02.¹³
 - (e) Annealed (at 1010°C for class 2 material and at 925°C for class 3 material; slow furnace cooling).^{14,15}
 - (f) Solution heat treated (at 1010°C for class 2 material and at 925°C for class 3 material; rapid water cooling), cold worked to ½ hard and precipitation hardened (at 465°C for 4.5h for class 2 material and at 450°C for 4h for class 3 material).^{14,15}
 - (g) Solution heat treated (at 1010°C for class 2 material and at 925°C for class 3 material; rapid water cooling) and precipitation hardened (at 465°C for 4.5h for class 2 material and at 450°C for 4h for class 3 material).^{14,15}
-

Out of the five RSW parameters that are controlled, the electrode force is not regarded as a predictive parameter because its value is kept constant for all RSW joints. Hence, each of the considered classifiers predicts the quality level of RSW joints from four RSW parameters: (i) welding time; (ii) welding current; (iii) electrode material; and (iv) treatment applied to electrode material.

2.3 Quality levels

The training of the predictive tools employs the 330 joints obtained by RSW process. For each of the RSW joints, the values of the four predictive parameters, which are shown in Table 2, and the quality level, assigned to the RSW joint by a human operator, are used. The quality level may be: (i) “acceptable”; or (ii) “unacceptable”.

The quality level is assessed by ultrasonic testing, except for the cases in which electrode-sheet sticking occurs (in 40 out of the 330 total RSW joints); in these cases the RSW joints are directly considered as “unacceptable” (Figure 1).

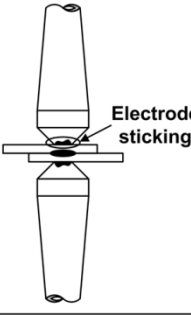
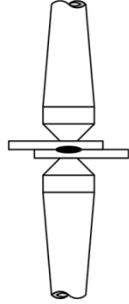
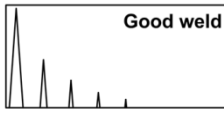
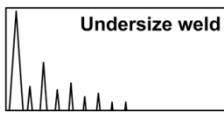
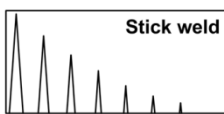
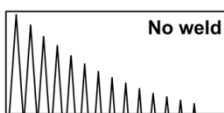
	Ultrasonic testing	Quality level
 <p>Electrode sticking</p>	Not applicable	Unacceptable
	 <p>Good weld</p>	Acceptable
	 <p>Undersize weld</p>	
	 <p>Stick weld</p>	Unacceptable
	 <p>No weld</p>	

Figure 1. Assessment of the quality level of the RSW joints by a human operator.

Since the weld nugget size is the most important parameter among those that determine the mechanical behaviour of the RSW joint,^{16,17} the quality level assessed by ultrasonic testing must be determined by the size of the weld nugget.¹⁸ The nugget is formed from the solidification of the molten metal and has a cast microstructure with coarse and columnar grains.^{4,19}

The human operator uses the ultrasonic testing to classify the 330 RSW joints into four categories (Figure 1) according to the effect of the weld nugget on the ultrasonic beam:^{4,19} (i) good weld (acceptable quality level): 123/330; (ii) undersize weld (unacceptable quality level): 86/330; (iii) stick weld (unacceptable quality level): 13/330; (iv) no weld (unacceptable quality level): 68/330.

2. 4 Computational Experiments

When a classifier is trained with a given data set, it often overfits the data, especially when using highly flexible models. This means that the classifier learns peculiarities of the training dataset which are not useful (and may even be detrimental) to predict in a general setting (i.e. when applied to another data set).

A sophisticated family of strategies to conduct model assessment (the evaluation of a model's performance) and model selection (selecting the model with the appropriate balance between flexibility and generalization ability) is cross-validation (CV). The k-fold cross-validation option involves randomly partitioning the original data set into k subsets of approximately equal size, called folds, and use k-1 of these subsets as training set and the remaining subset as independent test set. This process is then rotated k times and the results are averaged over the rounds.^{20,21} This approach is useful because it allows estimating the test error of the classifiers, computing additional measures about the dispersion of the performance and reducing the influence of a particular training/test division in the results. In practice, values of k=5 and k=10 have been shown to provide an adequate bias-variance trade-off without requiring excessive computational power.²¹

It is important to notice that even though CV can be used for model selection and model assessment, some caveats must be considered.²² Several studies^{23,24} have recently warned against using the error obtained in the selection phase as an estimate of the test error of the selected model (i.e. the error that the model will have on new data). Using several parametric configurations of a classifier and computing the error using CV is adequate to select the right parameters to use, but this error might be too optimistic if reported as an estimate of the performance of the selected model. The estimated errors of the different classifiers analysed in this work have been obtained using a CV scheme suggested as an unbiased estimate of the true performance error of the method. This approach is called nested cross-validation^{23,24}, although is also known with other names.²²

Nested CV uses two nested loops: the inner loop is used as model selection in which the parameters are estimated and tested without using all the available data; the outer loop employs the data that has not been used in the inner loop to compute an unbiased estimate of the performance of the model selected in the inner loop. Specifically, the inner loop uses as data the k-1 folds used as training in the outer loop. This data is in turn used in a k-fold analysis for every combination of parameters. For each fold in the outer loop, the model is trained with the combination of parameters with lower error in the inner fold.

In order to compare different classifiers, a performance evaluation metric is required. The most common measure used to assess the performance of a classifier is the percentage of correctly classified data, aka maximum classification rate or accuracy.²⁵ This metric is the proportion of correctly classified data instances in the test sets. Although informative, accuracy is not always the most appropriate measure for comparing classifiers. Among several problems,^{9,25,26} perhaps the most relevant in the context at hand is the implicit assumption of equal misclassification costs for false positives and false

negatives. When this is not appropriate, the objective function should be a misclassification cost which weighs false positives and false negatives differently.

In this work unacceptable RSW joints are considered positive cases and acceptable joints are considered negative cases. A true positive is an unacceptable joint predicted by the classifier as unacceptable, a true negative is an acceptable joint predicted as acceptable, a false negative is an unacceptable joint predicted as acceptable and a false positive is an acceptable joint predicted as unacceptable. Thus, a high rate of false negative joints can lead to safety problems while a high rate of false positives can increase the monetary costs of manufacturing and production unnecessarily.

Misclassification costs are often unknown or difficult to estimate, since they depend on the specific industrial objective and are often subject to change. Consequently, comparison among classifiers taking into account both types of errors is decomposed in two complementary performance measures, i.e. sensitivity (the proportion of positive samples which the classifier has correctly identified, eq. (1) and specificity (the fraction of negatives which the classifier has correctly identified, eq. (2), formally defined as:

$$Sensitivity = \frac{TP}{TP + FN} \quad (1)$$

$$Specificity = \frac{TN}{FP + TN} \quad (2)$$

Implicitly or explicitly, most classifiers contain a discrimination threshold (e.g. the minimum probability required to classify a sample as positive) that allows them to glide up and down the sensitivity-specificity trade-off. A common approach to assess binary classifiers accounting for this degree of freedom is the Receiver Operating Characteristic (ROC) curve.²⁷ The ROC curve of a classifier shows its true positive rate (or *sensitivity*) against its false positive rate (i.e. the fraction of misclassified negatives, aka fall-out, equal to $(1 - specificity)$) for different threshold values. The range of possible threshold values is chosen to include both the extreme classifier that classifies all data as negatives (false positive rate = 0, but true positive rate = 0) and the extreme classifier that classifies all data as positives (true positive rate = 1, but false positive rate = 1). The information contained in the ROC curve is often reduced to one single number by computing the Area Under the (ROC) Curve (AUC). The ideal classifier would have a true positive rate equal to one and a false positive rate equal to zero, so the larger the AUC, the better the classifier. An advantage of ROC curves is that they are insensitive to changes in class distribution; hence, this metric has no class skew.²⁷

2. 5 Classification Techniques

In this work we have compared the performance of seven classifiers, shown in Table 3 together with the R packages used in the computational experiments. A detailed description of the classification techniques and the parameters used can be found in Supplementary Material 1.

Table 3. Classification Algorithms and R packages used.

Classification Algorithm	Implementation details and R package
Tree	CART trees, “rpart” R package ²⁸
Pruned Tree	CART trees, “rpart” R package ²⁸
Boosting	“gbm” R package ²⁹
Neural Network	Resilient backpropagation with weight backtracking ³⁰ , “neuralnet” R package ³¹
Random Forest	“randomForest” R package ³²
SVM radial	“e1071” R package ³³
SVM linear	“e1071” R package ³³
Logistic Regression	“glm” R package
Quadratic Discriminant Analysis (QDA)	qda() function, “MASS” R package ³⁴

3 Results and discussion

The predictive power of the classifiers has been compared from different perspectives. Nested 10-fold cross-validation has been used to compare the accuracy of the different methods. Table 4 shows the average maximum classification rate of each classifier with its standard error.

Analysis of Variance (ANOVA) is used to test the differences in classifier performance. Note that although the accuracy results depicted in Table 4 seem to show differences between classifier performances, these ones could be due to randomness in the training and cross-validation data sets. Table 5 shows the ANOVA results, and confirms the statistical difference between classifiers.

In order to figure out which classifiers are significantly different from each other, Duncan’s multiple range test,³⁵ popular in Machine Learning,⁹ is applied. Apart from using a popular stepwise test such as Duncan’s multiple range, we have complemented the comparison using a test that follows a different approach

(Bayesian) to determine the significance of the differences: the Waller-Duncan k-ratio test.³⁶ This method has good properties from both the Bayesian and frequentist points of view.³⁷ Table 4 shows the classifier performance accuracy sorted in descending order, and grouped into subsets in which performance differences are not significant. Results obtained using Waller-Duncan k-ratio test show that the accuracy obtained in previous research using ANN^{8,19} is outperformed by Random Forests. However, results from Duncan's multiple range test are not discriminating enough, and the most significant conclusion is that Support Vector Machines using linear kernels exhibit the worst performance. Clearer differences among classifiers are found using AUC performance (Table 8).

Table 4. Duncan's multiple range test and Waller-Duncan's multiple range test for accuracy (alpha: 0.05). Classifiers with one or more letters in common are not significantly different.

Classifier	Accuracy	SE	Duncan Subgroup	Waller-Duncan Subgroup
RF	0.9576	0.0103	a	a
SVMr	0.9394	0.009035	ab	ab
Boosting	0.9273	0.01979	ab	abc
Logistic	0.9121	0.01313	ab	bc
ANN	0.9121	0.01657	ab	bc
PTrees	0.903	0.01906	b	bc
Trees	0.8939	0.01982	b	c
QDA	0.8455	0.01717	c	d
SVMI	0.797	0.02023	d	e

Table 5. ANOVA of accuracy classifier performance

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
classifier	8	0.195	0.02441	8.83	1.3e-08 ***
Residuals	81	0.224	0.00276		

Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

As previously discussed, accuracy is not an appropriate performance measure when the cost of a prediction failure –false negative or false positive– is not equivalent. Consequently the analysis has been complemented comparing the techniques in a wider range of situations using ROC curves and AUC performance. Calculations have been performed with the R package “ROCR”, from Sing *et al.*³⁸ using again a nested 10-fold cross-validation for each classifier. Figure 2 shows the mean ROC curve obtained averaging the 10 folds. Curves intersect, indicating the absence of a dominant classifier for every situation. For instance, Figure 3 shows that the sensitivity of a gradient tree boosting classifier for a fixed specificity of 75% is similar to a neural network classifier, but is higher when the specificity is fixed to 90% (numeric results can be found in Table 6).

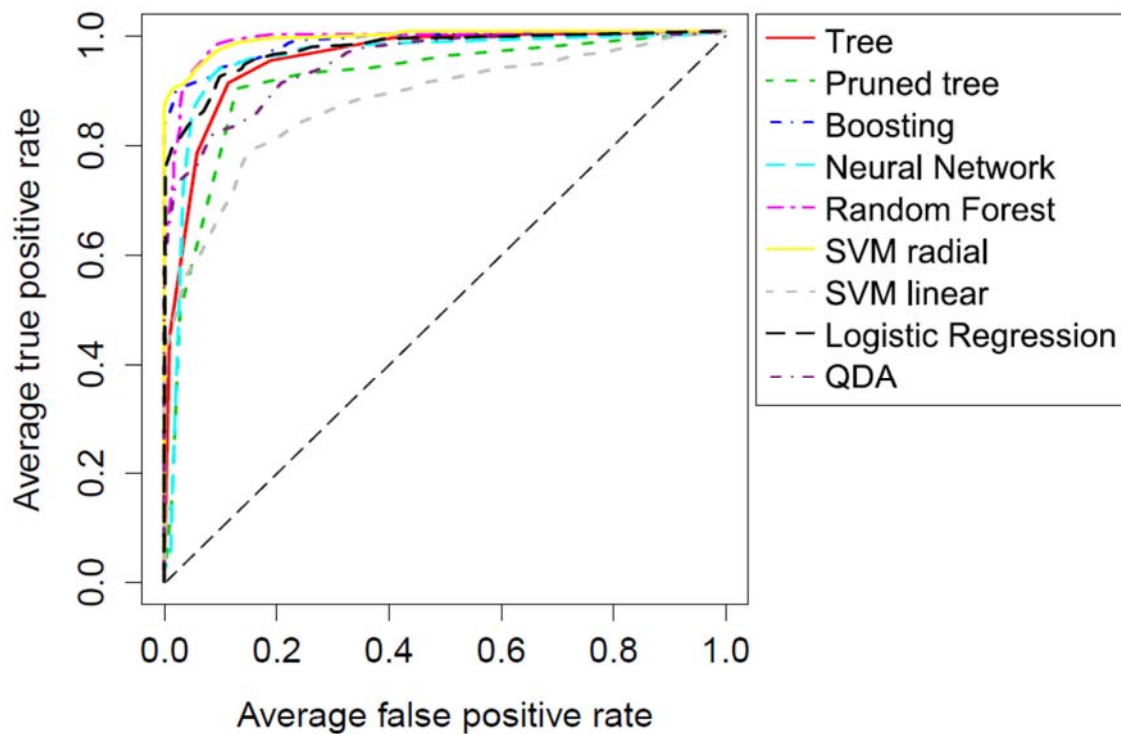


Figure 2. ROC curve for each tested classifier. Dotted diagonal represents random guessing.

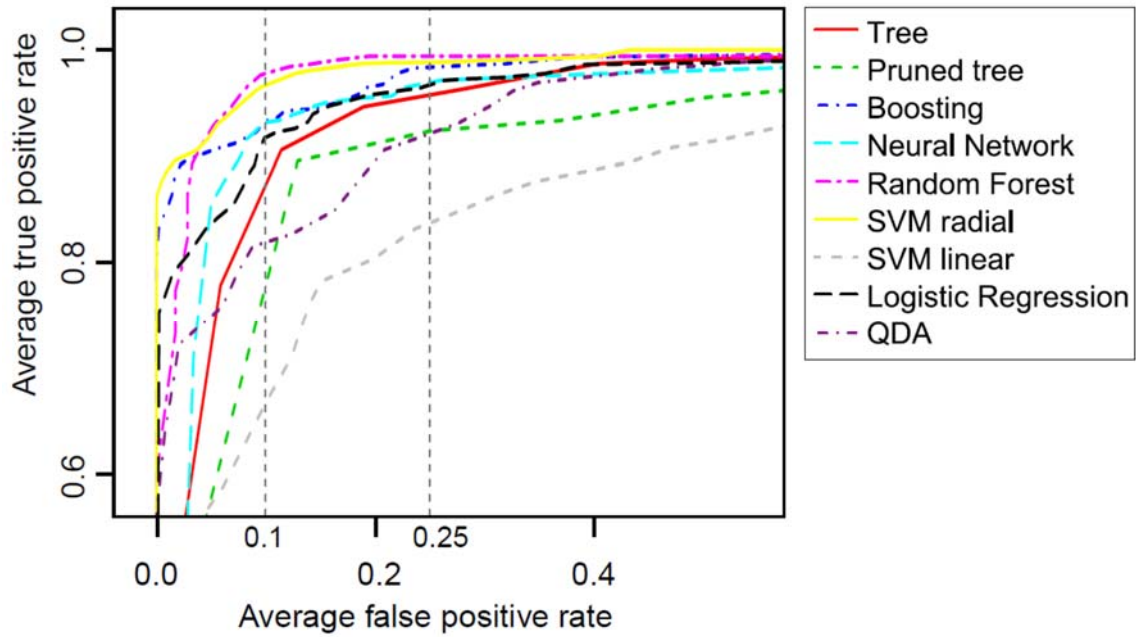


Figure 3. Enlargement of the plot in Figure 2 to show ROC curve for each tested classifier and sensitivity when a specificity of 75% or a specificity of 90% is required.

summarizes the average AUC, standard error and the sensitivity performance of the classifiers when a level of specificity is fixed.

Table 6. Average AUC, standard error and the sensitivity performance of the classifiers when the level of specificity is fixed to 75% or 90%.

Technique	Area Under ROC Curve		Sensitivity	
	Average AUC	Standard Error	Specificity of 75%	Specificity of 90%
Tree	0.9563	0.01274	0.9596	0.8745
Pruned Tree	0.9174	0.01888	0.9212	0.7757
Boosting	0.9805	0.0094	0.9816	0.9322
Neural Network	0.9448	0.01678	0.9706	0.9322
Random Forest	0.9831	0.00631	0.9953	0.9788
SVM radial	0.9899	0.00352	0.9898	0.9651
SVM linear	0.8773	0.01698	0.8361	0.6686

Logistic Regression	0.9735	0.00867	0.9706	0.9157
QDA	0.9396	0.01217	0.9212	0.8168

Again, results of classifiers using AUC performance are different. An ANOVA analysis has been conducted to test differences in performance (Table 7). Differences are again significant.

Table 7. ANOVA of AUC classifier performance

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
classifier	8	0.107	0.01333	8.26	4.1e-08 ***
Residuals	81	0.131	0.00161		

Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Support Vector Machines using radial kernel, Random Forests and Boosting obtain again the best overall results (Table 8). Waller-Duncan k-ratio test shows that the differences with ANN or QDA are statistically significant, however the differences obtained with Logistic regression or Trees are not enough to be considered significant with any of the used tests.

Table 8. Duncan's multiple range test and Waller-Duncan's multiple range test for AUC (alpha: 0.05). Classifiers with one or more letters in common are not significantly different.

Classifier	AUC	SE	Duncan Subgroup	Waller-Duncan Subgroup
SVMr	0.9899	0.00352	a	a
RF	0.9831	0.00631	a	a
Boosting	0.9805	0.0094	a	a
Logistic	0.9735	0.00867	ab	ab
Trees	0.9563	0.01274	ab	abc
ANN	0.9448	0.01678	ab	bcd
QDA	0.9396	0.01217	ab	cd
PTrees	0.9174	0.01888	bc	d
SVMI	0.8773	0.01698	c	e

In order to better interpret the effect of the different welding variables in the quality of the RSW joints, the different decision boundaries proposed by the best classifiers –boosting, random forests, and SVM radial– are compared in Figure 4. An analysis quantifying the relative effect of each one of the variables considered for classification can be found in supplementary material 2, and a multiclass analysis estimating the types of errors depending on the region can be found in supplementary material 3.

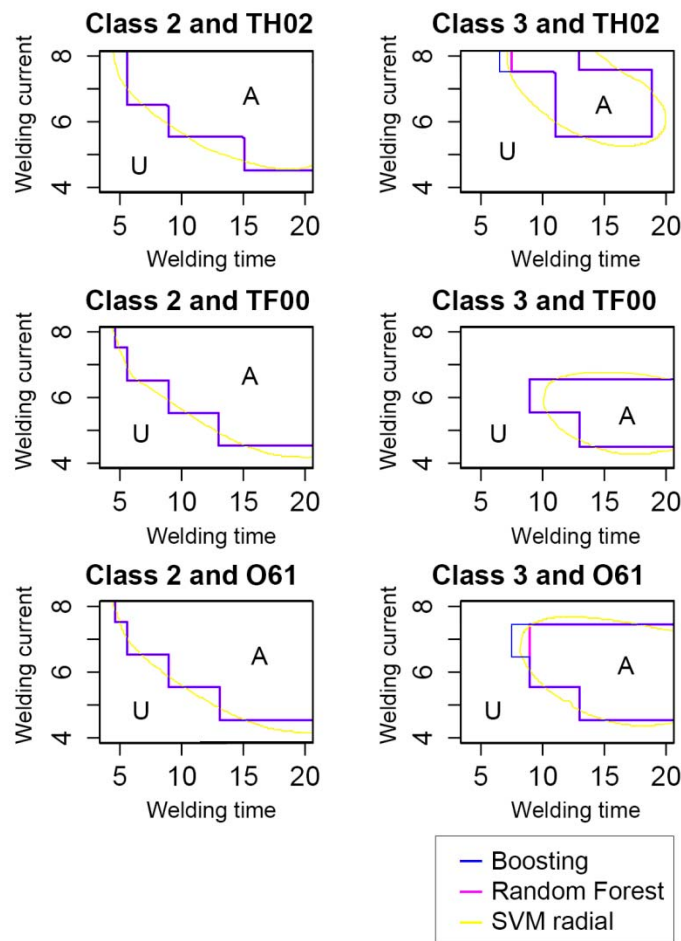


Figure 4. Decision boundaries computed by Boosting, Random Forest and SVM radial algorithms. “A” represents acceptable region, and “U” represents unacceptable region. Welding current, welding time and electrode effects are considered.

Class 2 electrode is less sensitive to treatment effects and offers better results than class 3 electrode since it has higher thermal and electrical conductivities.¹¹ On the other hand, class 3 electrode offers better mechanical properties¹¹ that prevent the electrode tip deformation (associated with the increase of the electrode contact face and the consequent current density decrease), which

occurs after a prolonged and continuous use. Nevertheless, this better behaviour of class 3 is not shown in results because, in this work, the electrodes are subject to less demanding conditions than those used in automotive industry when high productivity is sought.

In RSW processes, electrodes must have a good combination of: (i) sufficiently high thermal and electrical conductivities (to prevent electrode-sheet sticking); and (ii) adequate strength (to avoid deformation at high pressures and temperatures).¹¹ TF00 temper can achieve this good combination of physical and mechanical properties because the formation and growth of the strengthening precipitates also reduce the contents of solute atom in matrix and, hence, the electrical conductivity increases.^{39,40} This effect is accentuated in TH02 temper where the cold work prior to precipitation hardening gives rise to dislocations that provide additional nucleation sites at which heterogeneous nucleation can occur. The consequent increase of the density of precipitates for a given time of aging,^{41,42} not only enhances the strengthening but also increase the electrical conductivity. O61 temper, although leads to low mechanical properties, offers good results because, in the present work, the electrodes are not subject to demanding conditions that may cause the electrode tip deformation and the consequent current density decrease.

4 Conclusions

In this work some of the most relevant and popular pattern recognition techniques have been compared for classification of RSW joints using the welding parameters as inputs. The major conclusions are:

- (1) The analysis confirms that knowing the welding time, the welding current and the type of electrode (electrode material and treatment) is sufficient to obtain classification rates almost comparable with those obtained using non-destructive testing. Additional causes of disturbance can appear during and after the welding process (e.g. electrode degradation, expulsion, current shunting, greasy surface...) but the use of a direct controlling process can reduce significantly the workload of a subsequent quality control. The proposed methodology could be used to implement an anomaly detection algorithm that can warn in real time about potentially detrimental drifts in the welding process. Thus, problematic welding parametric regions could be detected before unacceptable RSW joints appear as a consequence of the direct setup.
- (2) According to Waller-Duncan k-ratio test, random forests significantly improve the classification performance (accuracy) obtained by previous research.⁸ These results suggest their use as effective decision support tools to assist directly in quality control of the RSW process, reducing post-

welding testing. The differences among random forests, support vector machines using radial kernel and boosting are not found significant.

- (3) Results show that for this problem there is not a dominant classifier for every possible pair specificity/sensitivity. An algorithm can perform better than others depending on the industrial context that determines the different cost of a prediction error. Notwithstanding, in an aggregated way, the analysis of the AUC performance measure shows that support vector machines using radial kernel, boosting, random forest and logistic regression using cubic terms or even decision trees are better candidates.

Acknowledgements

The authors would like to thank Dr. Luis R. Izquierdo for some advice and comments on this paper. The authors acknowledge support from the Spanish MICINN Project CSD2010-00034 (SimulPast CONSOLIDER-INGENIO 2010) and by the Junta de Castilla y León GREX251-2009.

References

1. J. H. Kim, Y. Cho and Y. H. Jang. 'Estimation of the weldability of single-sided resistance spot welding', *J. Manuf. Syst.*, 2013; 32, 505-512. DOI: 10.1016/j.jmsy.2013.04.007
2. O. Martín, M. López, P. De Tiedra and M. S. Juan. 'Prediction of magnetic interference from resistance spot welding processes on implantable cardioverter-defibrillators', *J. Mater. Process. Technol.*, 2008; 206, 256-262. DOI: j.jmatprotec.2007.12.021
3. S. M. Hamidinejad, F. Kolahan and A. H. Kokabi. 'The modeling and process analysis of resistance spot welding on galvanized steel sheets used in car body manufacturing', *Mater. Des.*, 2012; 34, 759-767. DOI: 10.1016/j.matdes.2011.06.064
4. Ó. Martín, M. Pereda, J. I. Santos and J. M. Galán. 'Assessment of resistance spot welding quality based on ultrasonic testing and tree-based techniques', *J. Mater. Process. Technol.*, 2014; 214, 2478-2487. DOI: 10.1016/j.jmatprotec.2014.05.021
5. H. J. Zhang, F. J. Wang, W. G. Gao and Y. Y. Hou. 'Quality assessment for resistance spot welding based on binary image of electrode displacement signal and probabilistic neural network', *Sci. Technol. Weld. Join.*, 2014; 19, 242-249. DOI: 10.1179/1362171813Y.0000000187
6. B. Parida and S. Pal. 'Fuzzy assisted grey Taguchi approach for optimisation of multiple weld quality properties in friction stir welding

- process', *Sci. Technol. Weld. Join.*, 2014; 20, 35-41. DOI: 10.1179/1362171814Y.0000000251
7. Y. R. Wong and S. F. Ling. 'Novel classification method of metal transfer modes in gas metal arc welding by real time input electrical impedance', *Sci. Technol. Weld. Join.*, 2013; 19, 224-230. DOI: 10.1179/1362171813Y.0000000184
 8. O. Martín, M. López and F. Martín. 'Redes neuronales artificiales para la predicción de la calidad en soldadura por resistencia por puntos', *Revista de Metalurgia.*, 2006; 42, 345-353. DOI: 10.3989/revmetalm.2006.v42.i5.32
 9. A. P. Bradley. 'The use of the area under the ROC curve in the evaluation of machine learning algorithms', *Pattern Recognit.*, 1997; 30, 1145-1159. DOI: 10.1016/S0031-3203(96)00142-2
 10. S. Simoncic and P. Podržaj. 'Resistance spot weld strength estimation based on electrode tip displacement/velocity curve obtained by image processing', *Sci. Technol. Weld. Join.*, 2014; 19, 468-475. DOI: 10.1179/1362171814Y.0000000212
 11. R. B. McCauley, M. P. Bennett, W. D. Bodary, G. C. Farrington, R. J. Gasser, W. W. Hurd, A. W. Schueler, T. W. Shearer and J. B. Silverberg. 'Resistance Spot Welding', in 'Metals Handbook, vol. 6: Welding and Brazing', (ed. T. Lyman), 401-424; 1971, Metals Park, OH, American Society for Metals.
 12. P. Podržaj and S. Simoncic. 'Resistance spot welding control based on the temperature measurement', *Sci. Technol. Weld. Join.*, 2013; 18, 551-557. DOI: 10.1179/1362171813Y.0000000131
 13. ASTM B 601-02. 'Standard Classification for Temper Designations for Copper and Copper Alloys-Wrought and Cast'. 2002, PA, ASTM International.
 14. J. C. Harkness and A. Guha. 'Beryllium-Copper and Beryllium-Nickel Alloys', in 'Metals Handbook Ninth Edition, vol. 9: Metallography and Microstructures', (ed. K. Mills et al.), 392-398; 1985, Metals Park, OH, American Society for Metals.
 15. G. Joseph and K. J. A. Kundig. 'Copper: Its Trade, Manufacture, Use and Environmental Status'. 1999, Materials Park, OH, ASM International.
 16. H. Moshayedi and I. Sattari-Far. 'Numerical and experimental study of nugget size growth in resistance spot welding of austenitic stainless steels', *J. Mater. Process. Technol.*, 2012; 212, 347-354. DOI: 10.1016/j.jmatprotec.2011.09.004
 17. Y. Luo, J. L. Li and W. Wu. 'Nugget quality prediction of resistance spot welding on aluminium alloy based on structureborne acoustic emission

- signals', *Sci. Technol. Weld. Join.*, 2013; 18, 301-306. DOI: 10.1179/1362171812Y.0000000102
18. T. Mansour. 'Ultrasonic testing of spot welds in thin gage steel', in 'Nondestructive Testing Handbook. Vol. 7: Ultrasonic Testing', (ed. P. McIntire), 557-568; 1991, Metals Park, OH, American Society for Nondestructive Testing.
 19. O. Martín, M. López and F. Martín. 'Artificial neural networks for quality control by ultrasonic testing in resistance spot welding', *J. Mater. Process. Technol.*, 2007; 183, 226-233. DOI: 10.1016/j.jmatprotec.2006.10.011
 20. C. Shao, K. Paynabar, T. H. Kim, J. Jin, S. J. Hu, J. P. Spicer, H. Wang and J. A. Abell. 'Feature selection for manufacturing process monitoring using cross-validation', *J. Manuf. Syst.*, 2013; 32, 550-555. DOI: 10.1016/j.jmsy.2013.05.006
 21. T. Hastie, R. Tibshirani and J. H. Friedman. 'The Elements of Statistical Learning: Data Mining, Inference, and Prediction'. 2009, New York, NY, Springer.
 22. D. Krstajic, L. Buturovic, D. Leahy and S. Thomas. 'Cross-validation pitfalls when selecting and assessing regression and classification models', *J. Cheminform.*, 2014; 6, 10. DOI: 10.1186/1758-2946-6-10
 23. S. Varma and R. Simon. 'Bias in error estimation when using cross-validation for model selection', *BMC Bioinform.*, 2006; 7, 91. DOI: 10.1186/1471-2105-7-91
 24. E. Anderssen, K. Dyrstad, F. Westad and H. Martens. 'Reducing over-optimism in variable selection by cross-model validation', *Chemometr. Intell. Lab. Syst.*, 2006; 84, 69-74. DOI: 10.1016/j.chemolab.2006.04.021
 25. S. Viaene, R. A. Derrig, B. Baesens and G. Dedene. 'A Comparison of State-of-the-Art Classification Techniques for Expert Automobile Insurance Claim Fraud Detection', *J. Risk Insurance.*, 2002; 69, 373-421. DOI: 10.1111/1539-6975.00023
 26. F. Provost and T. Fawcett. 'Robust Classification for Imprecise Environments', *Mach. Learn.*, 2001; 42, 203-231. DOI: 10.1023/A:1007601015854
 27. T. Fawcett. 'An introduction to ROC analysis', *Pattern Recognit. Lett.*, 2006; 27, 861-874. DOI: 10.1016/j.patrec.2005.10.010
 28. T. M. Therneau and E. J. Atkinson. 'An introduction to recursive partitioning using the rpart routines', in 'Division of Biostatistics 61', 1997, Mayo Foundation.
 29. G. Ridgeway. 'Generalized boosted regression models', *Documentation on the R Package 'gbm'*, version 1.5.7., 2006.

30. M. Riedmiller. 'Rprop - Description and Implementation Details. Technical Report.'. 1994, University of Karlsruhe.
31. F. Günther and S. Fritsch. 'neuralnet: Training of neural networks', *The R journal.*, 2010; 2, 30-37.
32. A. Liaw and M. Wiener. 'Classification and Regression by randomForest', *R News.*, 2002; 2, 118-122.
33. E. Dimitriadou, K. Hornik, F. Leisch, D. Meyer and A. Weingessel. 'Misc functions of the Department of Statistics (e1071), TU Wien. R package version 1.6-3'. 2008.
34. W. N. Venables and B. D. Ripley. 'Modern applied statistics with S'. 2002, New York, Springer.
35. D. B. Duncan. 'Multiple range and multiple F tests', *Biometrics.*, 1955; 11, 1-41. DOI: 10.2307/3001478
36. R. A. Waller and D. B. Duncan. 'A Bayes Rule for the Symmetric Multiple Comparisons Problems', *J. Am. Stat. Assoc.*, 1969; 64, 1484-1503.
37. J. P. Shaffer. 'A semi-Bayesian study of Duncan's Bayesian multiple comparison procedure', *J. Stat. Plan. Inference.*, 1999; 82, 197-213. DOI: 10.1016/S0378-3758(99)00042-7
38. T. Sing, O. Sander, N. Beerenwinkel and T. Lengauer. 'ROCR: visualizing classifier performance in R', *Bioinform.*, 2005; 21, 3940-3941. DOI: 10.1093/bioinformatics/bti623
39. N. Gao, T. Tiainen, Y. Ji and L. Laakso. 'Control of microstructures and properties of a phosphorus-containing Cu-0.6 Wt.% Cr alloy through precipitation treatment', *J. Mater. Eng. Perform.*, 2000; 9, 623-629. DOI: 10.1361/105994900770345476
40. J. h. Su, Q. m. Dong, P. Liu, H. j. Li and B. x. Kang. 'Research on aging precipitation in a Cu–Cr–Zr–Mg alloy', *Mater. Sci. Eng.: A.*, 2005; 392, 422-426. DOI: 10.1016/j.msea.2004.09.041
41. J. B. Clark. 'Age hardening in a Mg-9 wt.% Al alloy', *J. Artic. Acta Metallurg.*, 1968; 16, 141-152. DOI: 10.1016/0001-6160(68)90109-0
42. S. P. Ringer, B. C. Muddle and I. J. Polmear. 'Effects of cold work on precipitation in Al-Cu-Mg-(Ag) and Al-Cu-Li-(Mg-Ag) alloys', *Metallurg. Mater. Trans. A.*, 1995; 26, 1659-1671.