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 Please do not cite this document, use instead:

O. Martín, M. Pereda, J.I. Santos, J.M. Galán. Assessment of resistance spot welding quality based on ultrasonic testing and tree-based techniques, Journal of Materials Processing Technology, Volume 214, Issue 11, November 2014, Pages 2478-2487, ISSN 0924-0136. doi:10.1016/j.jmatprotec.2014.05.021
http://www.sciencedirect.com/science/article/pii/S0924013614001964

Assessment of resistance spot welding quality based on ultrasonic testing and tree-based techniques

Óscar Martín[a,*], María Pereda[b], José Ignacio Santos[b], José Manuel Galán[b]

[a] 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.

[b] 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.

* Corresponding author. Ph.: +34-983423533, Fax: +34-983423310, E-mail: oml@eii.uva.es (O. Martín)

Abstract

Classification And Regression Tree (CART) and Random Forest techniques were proposed as pattern recognition tools for classification of ultrasonic oscillograms of Resistance Spot Welding (RSW) joints. The results showed that CART models produced an acceptable error rate with high interpretability. These features may be used to understand and control the decision processes, instruct other human operators, compare margins of safety or modify them depending on the criticality of the industrial process. Compared with CART trees, random forests reduced the error rate at the cost of decreasing decision interpretability. The use of the agreement of the forest was proposed as a measure to reduce the workload of human operators, who would only have to focus on the analysis of ultrasonic oscillograms that are difficult to interpret.

Keywords: Resistance Spot Welding; Non-Destructive Ultrasonic Testing; Random Forest Technique; CART Trees; Classification; Quality Control

PREPRINT. Please do not cite this document, use instead: O. Martín, M. Pereda, J.I. Santos, J.M. Galán. Assessment of resistance spot welding quality based on ultrasonic testing and tree-based techniques, Journal of Materials Processing Technology, Volume 214, Issue 11, November 2014, Pages 2478-2487, ISSN 0924-0136. doi:10.1016/j.jmatprotec.2014.05.021 http://www.sciencedirect.com/science/article/pii/S0924013614001964

1 Introduction

Resistance Spot Welding (RSW) is, as reported by Jou (2003), a manufacturing process widely used for joining sheet steel in the automotive industry, mainly because, as pointed out by Martín et al. (2008), its high speed and adaptability for automation render it suitable for mass production. Importantly, the number of RSW joints per vehicle is very high (3000–7000 according to Hamidinejad et al. (2012)) and there can be significant variability in the quality of each of them. This is partly due to the fact that RSW is a complicated process that involves interactions of electrical, thermal, mechanical and metallurgical phenomena, as indicated by Moshayedi and Sattari-Far (2012). Raoelison et al. (2012) stated that the quality of RSW joints is a function of the size of the weld nugget, which is formed from the solidification of the molten metal after a heating by Joule effect. Thus, quality control of the RSW joints has been a longstanding challenge which, as pointed out by Martín et al. (2006), would significantly benefit from the design of cost-effective decision support tools. Furthermore, the tendency in the highly competitive automobile industry is to reduce the number of RSW joints per vehicle; this fact reinforces the interest in a tool capable of assisting in the quality control of RSW joints because the fewer the RSW joints per vehicle, the stronger the requirements for each of them.

Chen et al. (2009) indicated that ultrasonic non-destructive testing is a promising technology for estimating the quality of the RSW joints in automotive industry. In the same line, Spinella et al. (2005) reported that this testing technique represents a cost-reduction opportunity in the quality control of RSW joints but they also noted that it requires that human operators have some degree of training. Nonetheless, Rupin et al. (2014) stated that a complex material microstructure may disturb the ultrasonic wave propagation and, consequently, the interpretation of data coming from ultrasonic testing of welds can be a challenge; Zhang (2011) and Thornton et al. (2012) agreed that ultrasonic A-scan depends heavily on the human operator's experience and Barrera et al. (2001) indicated that there may be problems in the assessment of ultrasonic oscillograms, either because the oscillograms are difficult to interpret, or because the task of repeatedly interpreting oscillograms for a long period of time may compromise the human operator's efficiency. Consequently, a tool that could effectively assist the human operator with the interpretation and classification of ultrasonic oscillograms would significantly improve the performance of the quality control conducted using ultrasonic non-destructive testing of RSW joints. Martín et al. (2007) developed a tool based on artificial neural networks (ANNs) for the classification of ultrasonic oscillograms with good results, but with the drawback that the ANNs are "black boxes", i.e. they lack explanatory power. The underlying knowledge captured by the network during its training is not transparent to the user; consequently, ANNs used as classifiers do not offer any interpretability of the results.

Two tools based on classification trees were proposed as decision support systems in welding quality control. Initially, Classification And Regression Trees (CART), which is one of the best known decision tree learning algorithms, was used. As a starting point, ultrasonic oscillograms obtained by ultrasonic non-destructive testing were classified by a human operator. This training dataset fed the algorithm and allowed approximating the decision rules used by the

human operator. Afterwards, an ensemble technique developed by Breiman (2001), namely random forests, was analysed in order to increase the classification power.

2 Experimental Procedure

2.1 Materials and equipment

The material welded by the RSW process was sheet steel, whose chemical composition and mechanical properties are respectively shown in Table 1 and Table 2. The thickness of the sheet steel was 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

Table 2. Mechanical properties of the sheet steel.

Yield strength (MPa)	Tensile strength (MPa)	Total elongation (%)	Hardness (HV)
192	301	40	104

The steel sheets were welded in a single-phase alternating current (AC) 50 Hz equipment by using water-cooled truncated cone RWMA Group A Class 2 electrodes with 16 mm body diameter and 5 mm face diameter.

The ultrasonic testing of the RSW joints employed a transducer that uses a captive water column delay and a replaceable rubber membrane with the aim of achieving good coupling with the surface of the RSW joint, which may have an indentation caused by the electrodes, as Fig. 1 shows. The frequency and the diameter of the transducer were respectively 20 MHz and 4.5 mm.

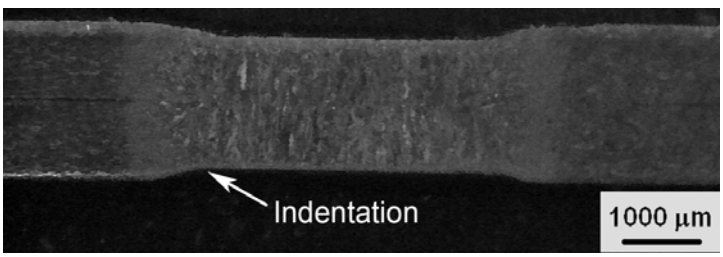

Fig. 1. Macrograph of a cross-sectioned RSW joint, whose contact surface with the electrodes has an indentation.

2.2 Welding conditions

A total of 438 joints were obtained by the RSW process. The controlled RSW parameters were welding time, welding current and electrode force which, as indicated by Aslanlar (2006), are the most important welding parameters in RSW. The values of these parameters are shown in Table 3. The values recommend in McCauley et al. (1971) were taken as reference but, whilst the electrode force was kept constant for all RSW joints, the welding time and the welding current took different values with the aim of achieving not only the quality level "good weld" but also the other three quality levels which were considered for classifying RSW joints and which will be explained later.

Table 3. RSW parameters and their respective values.

Welding time (s)	Welding current (kA RMS)	Electrode force (N)
0.08		
0.12		
0.16	4	
0.20	5	
0.24	6	980.7
0.28	7	
0.32	8	
0.36		
0.4		

2.3 Ultrasonic testing of RSW joints

The pulse-echo method with A-scan data presentation was used for ultrasonic testing. Zhang (2011) stated that the ultrasonic oscillogram obtained by the A-scan technique is a plot of wave amplitude versus time; the ultrasonic beam is reflected when it encounters an interface and then a series of echoes, which is associated with the quality of the RSW joint, is obtained: (i) the location of a reflecting interface is defined by the echo positions; and (ii) the height of the echoes depends on the sound attenuation which, in turn, depends on the microstructure of the weld nugget of the RSW joint.

2.3.1 Ultrasonic oscillograms data

A representative vector of each ultrasonic oscillogram must be obtained with the aim of processing the data obtained by ultrasonic testing. In order to be representative, a vector must contain information about the main factors that characterize an ultrasonic oscillogram. These factors are, according to Mansour (1991), the attenuation of the ultrasonic beam and the presence of one-layer echoes between principal echoes. It is also important that the vector does not contain redundant information. The criterion that determines the vector size is the number of echoes that are considered in the oscillogram: too few echoes give rise to vectors that may not be sufficiently representative, but too many echoes result in a vector with

redundant information. Martín (2004) developed a program that obtains a 10-component vector that uses only the first six echoes, where:

- The first five components of the vector are the relative heights of the echoes. Specifically, the n^{th} component (with n=1, . . ., 5) is the height of the $[n + 1]^{th}$ echo (h_{n+1}) divided by the height of the 1^{st} echo (h_1). Due to the role of the first echo in obtaining the vector, the ultrasonic testing equipment settings were adjusted so that the first echo displayed on the oscillogram was the first backwall echo.
- The last five components of the vector are the distances between consecutive echoes: the n^{th} component (with n=6, . . ., 10) is the distance between the $[n-4]^{th}$ echo and the $[n-5]^{th}$ echo (d_{n-5}).

2.3.2 RSW quality levels

The quality level of each RSW joint was determined from its ultrasonic oscillogram. According to Moshayedi and Sattari-Far (2012), weld nugget size is the most important parameter among those that determine the mechanical behaviour of the RSW joint; therefore, the quality level estimated by ultrasonic testing must be a function of the weld nugget size. Mansour (1991) indicated that, indeed, the oscillogram depends on the effect of weld nugget size on the ultrasonic beam for two reasons: (i) the weld nugget –which has melted and solidified– has a cast microstructure with coarse and columnar grains, as it is shown in Fig. 2; the weld nugget produces higher attenuation than the parent metal and, hence, the greater the thickness of the weld nugget, the higher the attenuation; (ii) the interface between the two sheets gives rise to the reflection of the ultrasonic beam; and one-layer echoes appear between principal echoes if the weld nugget diameter is smaller than the ultrasonic beam width.

Fig. 2. Micrograph of the weld nugget of a RSW joint, which has a cast microstructure with coarse and columnar grains.

According to Krautkrämer and Krautkrämer (1990), and taking into account the effect of weld nugget on the ultrasonic beam behaviour, four quality levels were considered for classifying RSW joints: (i) good weld; (ii) undersize weld; (iii) stick weld; (iv) no weld. The tensile-shear test was performed on the different quality levels according to ISO 14273:2000, and the results are provided in Fig. 3.

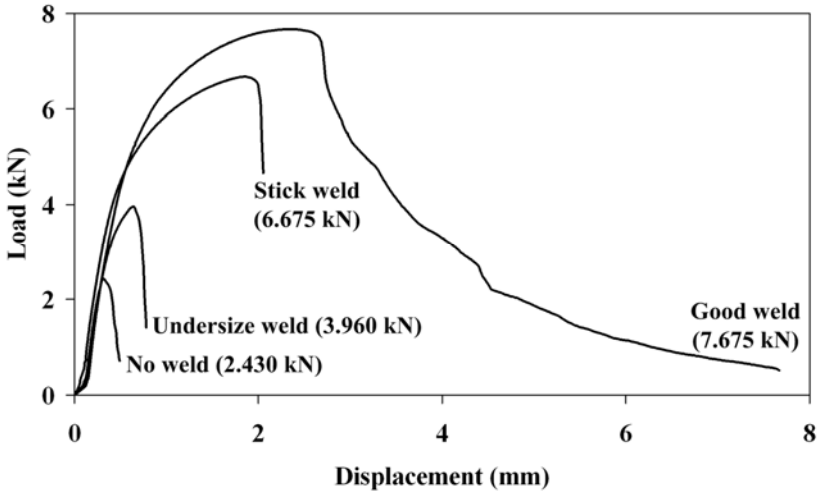

Fig. 3. Load vs. displacement curves and tensile shear load bearing capacities obtained from the tensile-shear test performed on the different quality levels.

2.3.2.1 Good weld

The weld nugget thickness is large and its diameter is greater than the ultrasonic beam width. Therefore, as Fig. 4 shows: (i) the span of the sequence of echoes is short due to the high attenuation caused by the coarse microstructure of a thick weld nugget; and (ii) the distance between consecutive echoes is the sum of the thickness of each of the two sheets because the ultrasonic beam reflections occur at the bottom surface of the lower sheet.

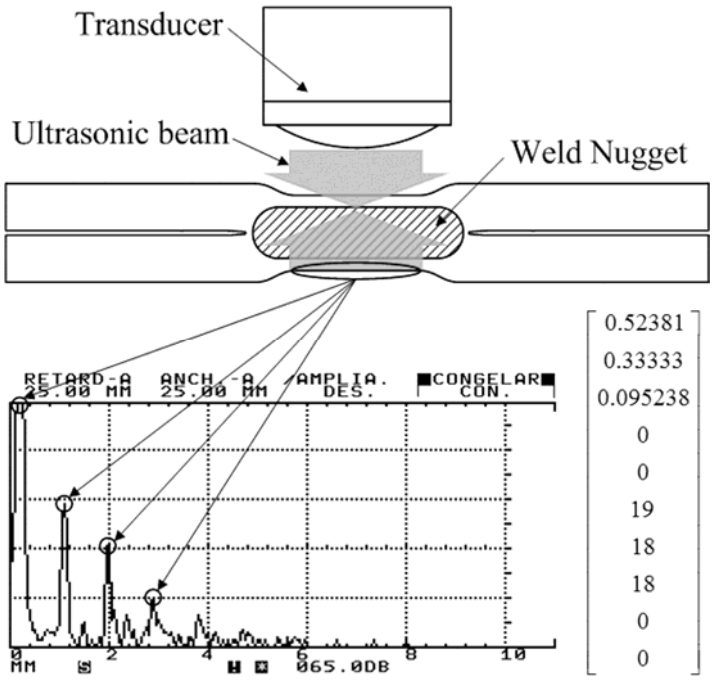

Fig. 4. Reflections of the ultrasonic beam for a good weld (top), ultrasonic oscillogram of a good weld (lower left) and its representative vector (lower right).

2.3.2.2 Undersize weld

Mansour (1991) indicated that the ultrasonic beam width must be approximately equal to the smallest allowable weld nugget diameter which, according to Kamiski (1997), should be at least four times the root of the thinner sheet thickness. Hence, the smallest allowable weld nugget diameter was set at 4.5 mm –a RSW joint was classified as undersize weld if its weld nugget diameter was smaller than 4.5 mm– and, consequently, the selected transducer diameter was 4.5 mm.

As it is shown in Fig. 5, since the weld nugget diameter is smaller than the ultrasonic beam width: (i) the part of the ultrasonic beam that passes through the weld nugget leads to principal echoes associated with the ultrasonic beam reflections which occur at the bottom surface of the lower sheet; whilst (ii) the part of the ultrasonic beam that does not pass through the weld nugget, and whose reflections occur at the interface between the two sheets, gives rise to *one-layer* echoes between principal echoes.

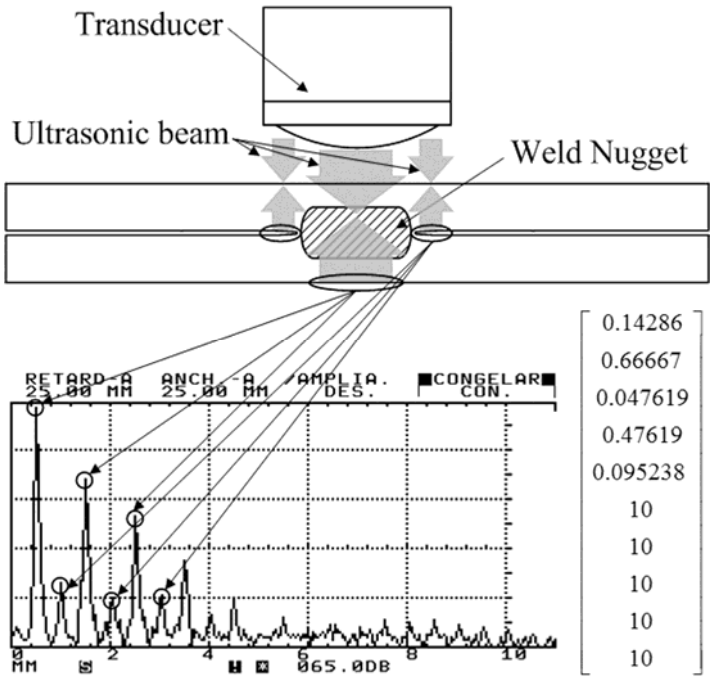

Fig. 5. Reflections of the ultrasonic beam for an undersize weld (top), ultrasonic oscillogram of an undersize weld (lower left) and its representative vector (lower right).

2.3.2.3 Stick weld

The criterion used for classifying a RSW joint as stick weld was based, as reported by Zhang (2011), on the heat input and on the attenuation that the weld nugget, melted by this heat input, causes in the ultrasonic beam. A stick weld is originated by a low heat input that gives rise to a weld nugget with an adequate diameter (i.e. larger than the ultrasonic beam width) but with a smaller thickness and –according to Lancaster (1999) and Özyürek (2008), which indicated that finer microstructures of the weld metal correspond with lower heat inputs– with a finer microstructure than that of a good weld. Hence, a stick weld, as it can be seen in Fig. 6: (i) produces lower attenuation than a good weld and the span of the sequence of echoes is longer; and (ii) the distance between consecutive echoes is the sum of the thickness of each of the two sheets because the ultrasonic beam reflections occur at the bottom surface of the lower sheet.

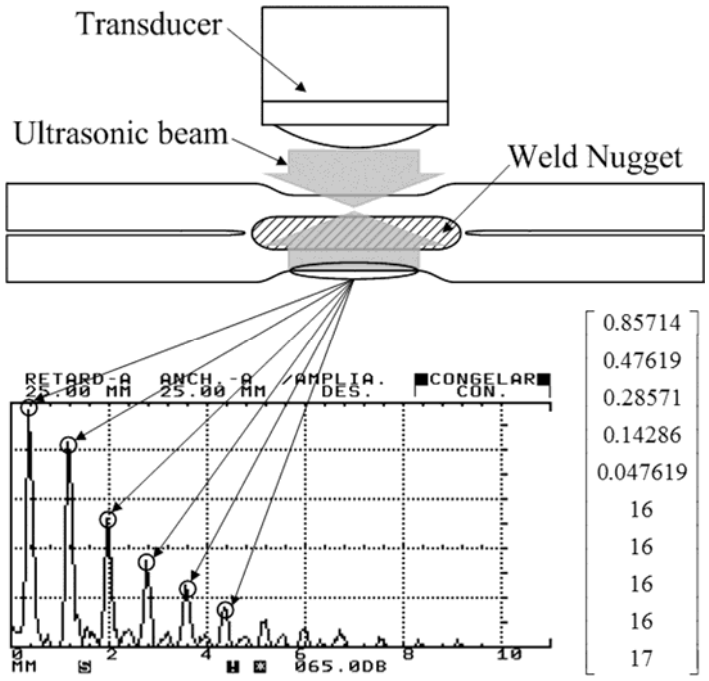

Fig. 6. Reflections of the ultrasonic beam for a stick weld (top), ultrasonic oscillogram of a stick weld (lower left) and its representative vector (lower right).

2.3.2.4 No weld

There is no melted and solidified metal; thus, the grain is not as coarse as the grain of a cast microstructure and the attenuation of the ultrasonic beam is lower than that of a RSW joint with weld nugget. Therefore, as it is shown in Fig. 7: (i) the span of the sequence of echoes is longer than that of a RSW joint with weld nugget; and (ii) the distance between echoes equals the thickness of one sheet because, since there is no weld nugget, there is no continuity between the two sheets.

PREPRINT. Please do not cite this document, use instead: O. Martín, M. Pereda, J.I. Santos, J.M. Galán. Assessment of resistance spot welding quality based on ultrasonic testing and tree-based techniques, Journal of Materials Processing Technology, Volume 214, Issue 11, November 2014, Pages 2478-2487, ISSN 0924-0136. doi:10.1016/j.jmatprotec.2014.05.021 http://www.sciencedirect.com/science/article/pii/S0924013614001964

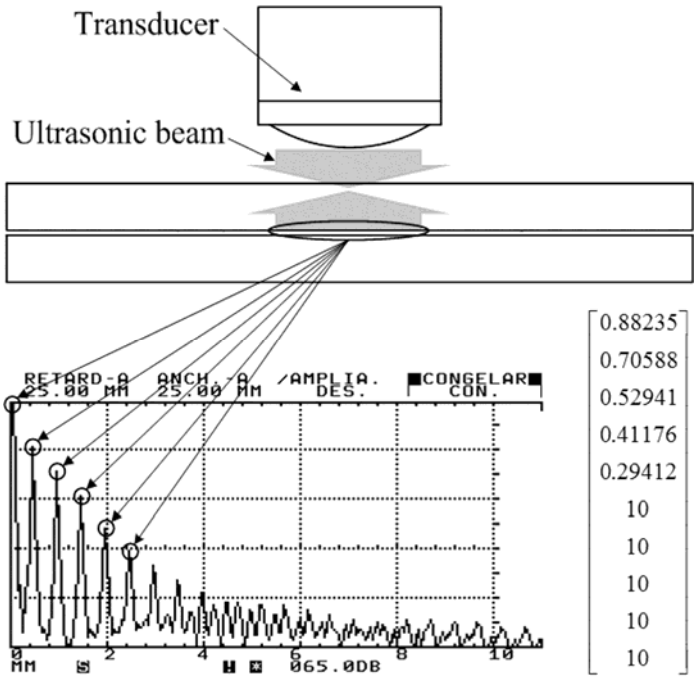

Fig. 7. Reflections of the ultrasonic beam for a no weld (top), ultrasonic oscillogram of a no weld (lower left) and its representative vector (lower right).

3 Theoretical background

A tree is a collection of nodes and links hierarchically organized in which every node –with the only exception of the root– has one incoming link. A decision tree is a non-parametric model typically used to make regression (i.e. quantitative prediction) or classification (i.e. qualitative prediction). When the tree is used for classifying, each node denotes a certain test. The input case to be classified starts at the root of the tree (i.e. the upper-most node), and goes down the tree following a path which is determined by the result of the tests conducted at each of the nodes in the way. The nodes in the path send the input to the appropriate leaf (child node) until the process reaches a terminal node. Terminal nodes usually associate the input with an output; this output may be a certain class label or a probability distribution over all possible classes.

There are several tree-based methods used for classification, although the most commonly used are C4.5 by Quinlan (1993) and CART trees by Breiman et al. (1984). In this work CART trees were initially used. Succinctly, a CART classification tree is built from a set of training or learning data. This set is a $N \times (M+1)$ matrix, where N denotes the number of RSW joints or examples of the set, M is the number of features or attributes x_j that defines each example, and the additional dimension includes the class $c \in C$ in which each case falls.

The tree is built top-down according to a splitting rule in each node. Here, binary trees with rules involving one single variable were considered. In such trees, a rule is just a question of the form: "$x_j > \theta$?", i.e. a test that checks whether the value of one specific attribute x_j of the input case is greater than a certain number θ, or not. Thus, the definition of a rule requires selecting both a certain feature x_j and a dividing value θ. The aim of the algorithm is to divide

the learning sample into subsets such that most observations within each subset belong to the same class (or as few classes as possible).

The process formally consists in selecting –recursively and greedily– the variable x_j –among the M attributes– and the value θ that best splits the set at each step according to a given metric. The metrics used to quantify how well classes have been separated are generically called impurity measures. As reported by Hastie et al. (2009) and Rokach and Maimon (2008), the two most popular impurity measures used to build trees are "information gain" and the "Gini index". The performance of both were analysed in this work.

Information gain is based on the concept of Shannon entropy. Entropy is a magnitude that tries to measure how mixed the data is with respect to a certain target variable (i.e. the class to which it belongs, in the present case). Thus, if all classes are equally represented in the dataset (maximal mixture), then the entropy is maximal; conversely, if all observations in the dataset belong to one single class (minimal mixture), the entropy of the dataset is 0. Therefore, for the present classification purposes the aim is to reduce the entropy as much as possible.

Specifically, given a set *S* of training examples that reach a node and its empirical distribution over classes *p(c)*, the Shannon entropy associated to the data S can be calculated as:

$$H(S) = -\sum_{c \in C} p(c) \cdot \log(p(c)) \tag{1}$$

where $p(c)$ is the proportion of input cases that are classified in class *c*. (In the case of $p(c)=0$ for any class, the value of $0 \cdot \log(0)$ is considered to be 0, consistently with the limit of the expression when p approaches 0.) After the node, two distributions S^1 and S^2, with their respective entropies $H(S^i)$, are obtained. Thus, the entropy after the split can be defined as the weighted average of the entropies of the resulting distributions S^1 and S^2, where the weights account for the number of cases in each of the distributions (see the last term in equation 2 below). Finally, the expected reduction in entropy –or information gain– can be defined as the difference between the entropy before the split, $H(S)$, and the entropy after the split:

$$I = H(S) - \sum_{i \in \{1,2\}} \frac{|S^i|}{|S|} H(S^i) \tag{2}$$

Naturally, the algorithm tries to find the rule that maximizes the information gain, or expected reduction in entropy.

In the case of using the Gini index function instead of using Eq. 1, the index of a set *S* can be calculated as:

$$G(S) = -\sum_{c \in C} p(c) \cdot (1 - p(c)) \tag{3}$$

The Gini impurity is a measure of how often a case would be misclassified if cases were stochastically assigned to classes according to the distribution of classes in the dataset. As with

entropy, the Gini impurity is a metric to be minimized, so an analogous optimization is performed.

Once an impurity measure is selected, the algorithm used for learning the tree considers every possible θ in the whole space of features of the training data set to find the optimal division. Every split separates the data in two leaves and the process is repeated recursively, finding the best split according to the impurity measure each time. Although, as pointed out by Criminisi et al. (2011), other splits different from linear functions in one variable ($x_j > \theta$) could be considered to train trees, they can increase the computational cost to find the optimum and hinder the interpretation of the decision rules of each node.

If the algorithm continues recursively, it builds a maximum tree with terminal nodes that contain observations of one class only. However, maximum trees suffer from a number of drawbacks: they are usually very complex and hence difficult to interpret, and when used with new data they do not work particularly well, since they often learn patterns that are specific to the training set but are not general for the problem, i.e. they "overfit". Consequently, maximum trees have to be optimized by cutting off irrelevant nodes and subtrees, i.e. they have to be "pruned". Among the different pruning strategies available, cost-complexity pruning using 10-fold cross validation has been used. This technique is based on the misclassification rate of the tree when applied to new data.

Classification trees built in this way show several desirable features, namely: they can effectively handle both continuous and discrete variables, they are non-parametric, invariant to monotonic transformations, robust with regard to outliers in learning data, and amenable to interpretation and understanding of the classification patterns. These advantages, together with reasonable classification performance in many contexts, justify their use in different fields such as industrial settings (applied by Georgilakis et al. (2007) for the selection of winding material in power transformers), medicine (used by Zhang and Li (2012) in the discrimination of melanoma from other benign affections), or ecological data analysis (see e.g. the use of trees by De'Ath and Fabricius (2000)). Notwithstanding, CART also presents some disadvantages, especially problems of high variance (i.e. tendency to overfit).

A powerful approach proposed by Breiman (1996) to reduce variance in machine learning, especially in high-variance, low-bias methods such as trees, is the use of bagging or bootstrap aggregation. Given a data set of N elements, the basic idea of bootstrapping involves creating m new datasets of the same size as the original one (i.e. N elements). Each of the new sets is obtained by sampling the original set with replacement. Bagging consists in fitting m unbiased but potentially noisy classification models, called weak learners, using the aforementioned bootstrap samples as training set, and combining them together to form a strong learner that is used to obtain a more robust classification (see Hastie et al. (2009) for more details).

Breiman (2001) adopted and modified the idea of bagging using partially uncorrelated trees to create *random forests*. The random forest algorithm to create a forest with *ntree* trees from a training set of N instances is as follows:

1) For t =1 to *ntree*

a. Bootstrap a sample of size N from the training set
b. Build a maximum (i.e. not pruned) tree T_t from the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size n_{min} is reached
    i. Select *mtry* variables at random from the M variables
    ii. Pick the best variable/split-point among the *mtry*
    iii. Split the node into two leaves with their corresponding nodes

2) Output the ensemble of *ntree* trees $\{T_t\}_1^{ntree}$

In order to classify a new datum x, let $\hat{C}_t(x)$ be the class prediction of the t-th tree of the random forest. Then the classification given by the random forest is obtained using the majority rule, i.e. every tree votes, and the classification provided is the one with the highest number of votes within the forest $\hat{C}_{rf}^{ntree}(x) = majority\ vote\left\{\hat{C}_t(x)\right\}_1^{ntree}$.

The reduction of variance combining trees via voting comes from the assumption that the trees are different from each other; naturally, the more similar the trees, the less advantageous is to combine them. Breiman (2001) obtained the needed de-correlation among trees injecting two sources of randomness in the process. First, each tree is built based on a different training dataset as a consequence of using the bootstrapping process. A second source of randomness is also added since splits are based on randomly selected attributes. This second process involves a trade-off. If *mtry* is close to M, the strength of each individual learner increases since they can operate using more information to classify; however, using high values of *mtry* also increases correlation among trees, thus reducing the advantage of combining weak learners. Brieman (2001) suggested testing with three possible values for *mtry* ($\frac{1}{2}\sqrt{M}$, $\sqrt{M}$, $2\sqrt{M}$) and choosing the one with the best performance, but the optimal balance may depend on the problem and should be determined through experimentation.

Random forests are fast and parallelizable, very resistant to overfitting, and they often obtain good results in different contexts. As indicated by Criminisi et al. (2011), all these features are contributing to their popularity. They also offer two additional relevant aspects based on the concept of "Out-Of Bag" (OOB) data. Note that each tree of the forest uses a different bootstrap sample from the original data set. About one third of the original data is not used in the construction of each tree, and these unused data is the so-called out of bag data. OOB data may be used as a test dataset to estimate misclassification error. The process is conducted classifying each datum only with those classifiers that have not been trained using such particular datum, and comparing them with its real class. OOB error is the proportion of misclassified data. Although still controversial, as pointed out by Mitchell (2011), OOB error has been proposed as an unbiased estimate of the true prediction error and, consequently, a cross-validation alternative for random forests. A second interesting analysis obtained from OOB data consists in determining the relative importance of each variable to classify data. For each tree, OOB sample is used to permute the j-th $\in M$ feature of these data at random. Then, OOB error of the tree is computed with these modified data. The importance of the j-th feature is the increase of OOB error obtained as a consequence of the permutation. The higher

the increase of the OOB error, the more important is the variable to achieve a correct classification.

In order to analyse and compare the predictive power of each technique and how the results generalize to an independent data set, cross-validation was used in each of the different approaches followed in this work. Cross-validation can be performed partitioning available data into two complementary sets, i.e. a set for training and a set for cross-validation. When the subsets do not contain a large number of instances, a strategy to reduce variability in cross-validation results is to perform multiple rounds of cross-validation using different partitions and to calculate the results averaging over the rounds. One approach to do this is the so-called k-fold cross-validation: the original data are randomly partitioned into k subsets of equal size, k-1 subsets are used to train and the remaining data is used to test. The process is rotated k times and results are averaged. The extreme case of k-fold validation is Leave-One-Out Cross-Validation (LOOCV), in which k is equal to the number of examples in the original data. This strategy implies using a single observation as validation set and the remaining data to train the model and repeat the process as many times as the number of original instances (see Fig. 8). LOOCV involves more computational calculations but implies training with more data.

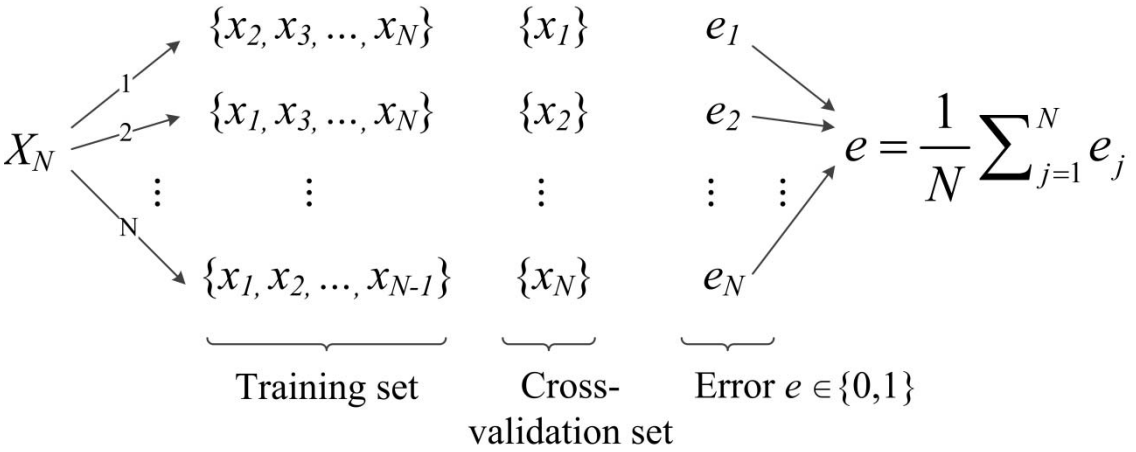

Fig. 8. Scheme of the LOOCV cross-validation technique. The expected level of accuracy of the model is measured by the average of the classification errors in all N rounds.

4 Results and discussion

4.1 Decision trees

The results of CART trees using maximum and pruned trees, and considering Gini index and information gain splitting rule strategies, were analysed. CART trees were implemented with the R package "rpart", developed by Therneau and Atkinson (1997), with the parameters Cp=0.001 and MinSplit=1. The pruned trees using the complete data set of 438 examples are presented in Fig. 9 for the case of information gain (Eq. 1) and in Fig. 10 for the Gini index (Eq. 2).

PREPRINT. Please do not cite this document, use instead: O. Martín, M. Pereda, J.I. Santos, J.M. Galán. Assessment of resistance spot welding quality based on ultrasonic testing and tree-based techniques, Journal of Materials Processing Technology, Volume 214, Issue 11, November 2014, Pages 2478-2487, ISSN 0924-0136. doi:10.1016/j.jmatprotec.2014.05.021 http://www.sciencedirect.com/science/article/pii/S0924013614001964

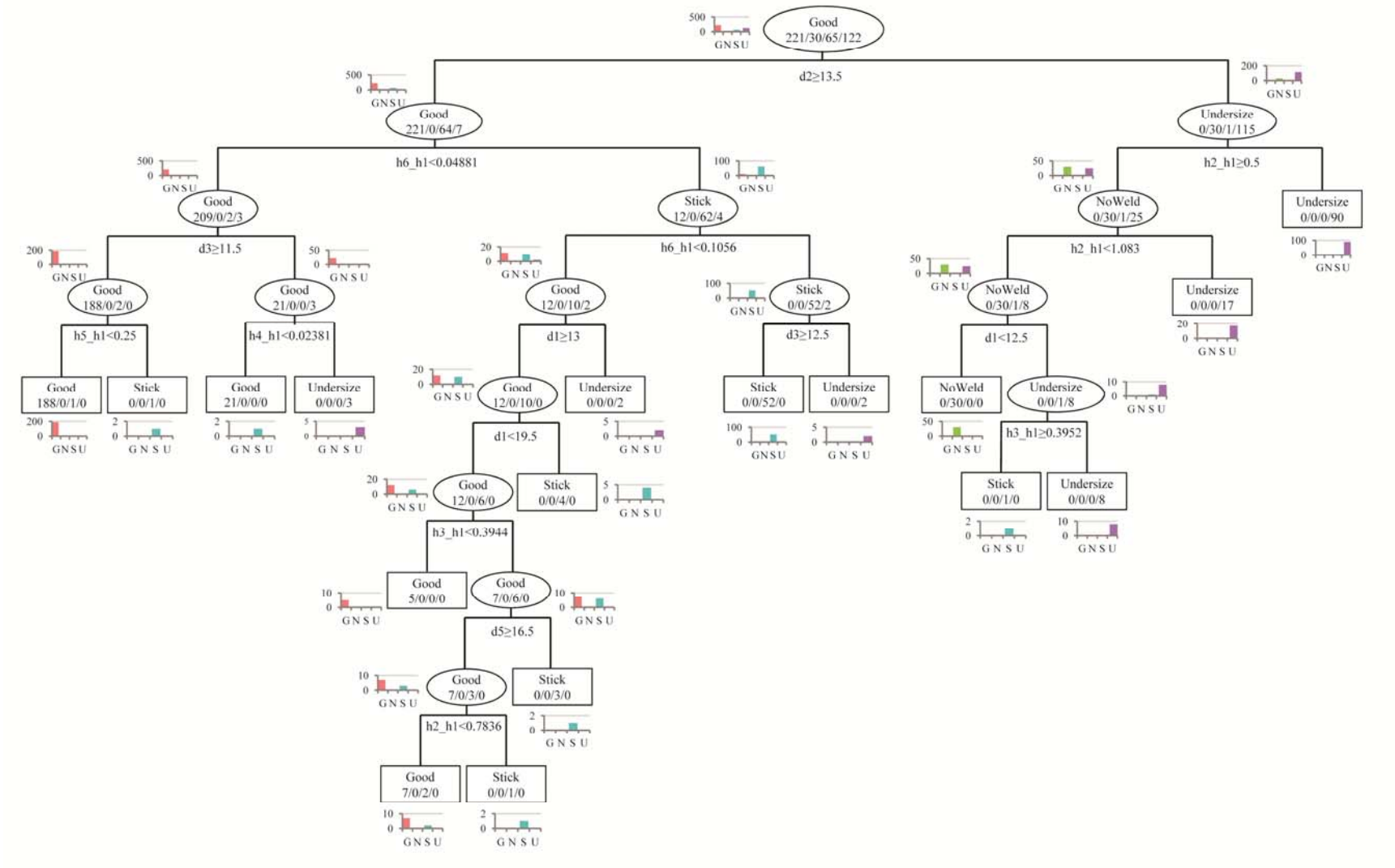

Fig. 9. Pruned classification tree using information gain strategy. Supplementary table 1 related to this figure can be found, in the online version, at

PREPRINT. Please do not cite this document, use instead: O. Martín, M. Pereda, J.I. Santos, J.M. Galán. Assessment of resistance spot welding quality based on ultrasonic testing and tree-based techniques, Journal of Materials Processing Technology, Volume 214, Issue 11, November 2014, Pages 2478-2487, ISSN 0924-0136. doi:10.1016/j.jmatprotec.2014.05.021 http://www.sciencedirect.com/science/article/pii/S0924013614001964

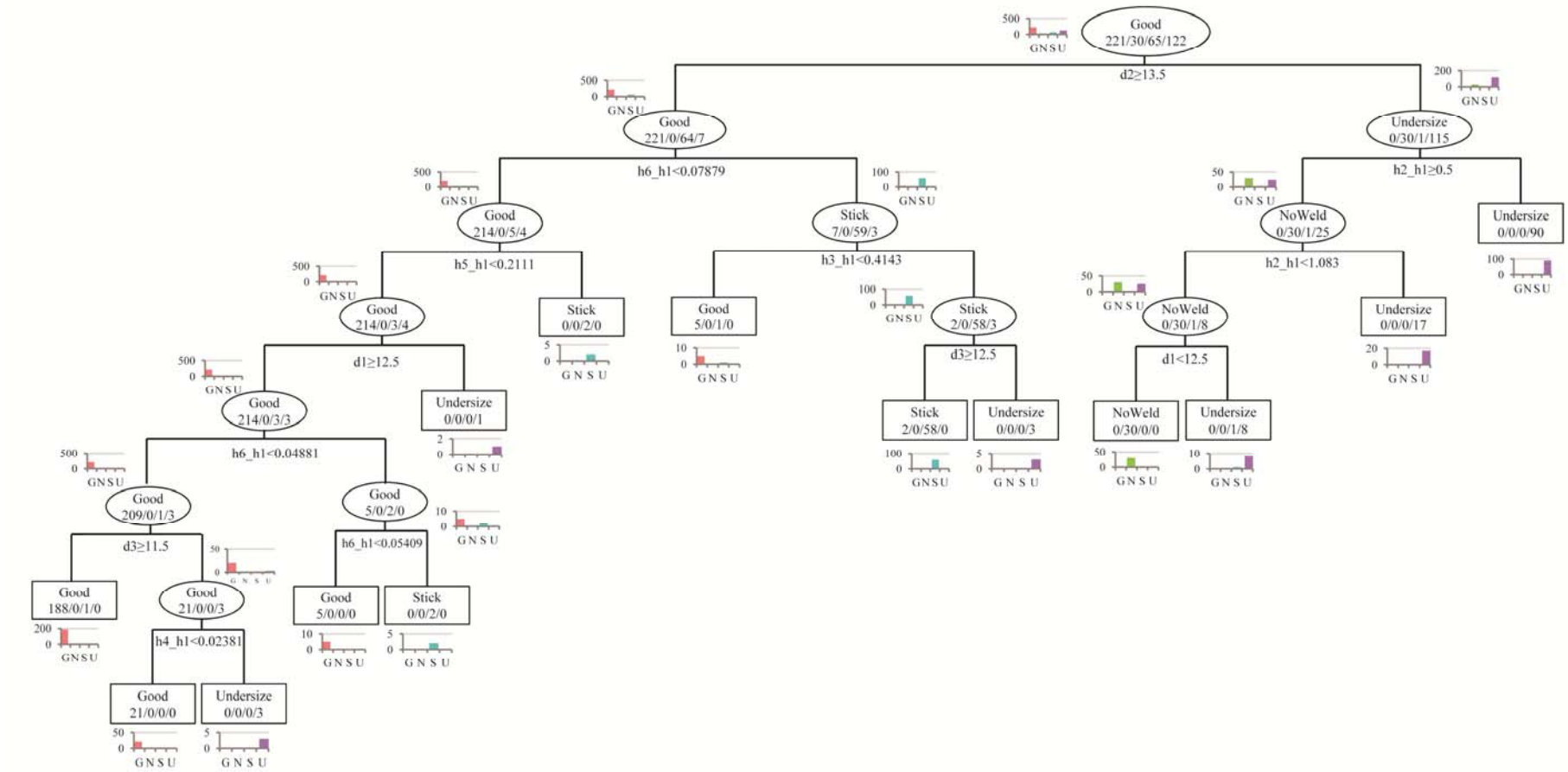

Fig. 10. Pruned classification tree using Gini strategy. Supplementary table 2 related to this figure can be found, in the online version, at

In the figures, the ovals denote decision nodes and the rectangles denote end nodes. The text label of the node shows the most frequent class of the data set that has reached each node. Under each decision node there is a condition that separates the data into two sets: on the left, the data that satisfies the condition, and on the right, the data that does not satisfy the condition. The histograms represent the number of data of each class of the training set that meet all conditions until this node. Each bar of the histograms represents a class, meaning G: Good, N: No Weld, S: Stick, U: Undersize.

As an example, the first three nodes of each decision tree in Figs. 9 and 10 are studied. The three features used are the following: (i) the distance between the second and third echoes (d_2), which allows knowing if there is continuity between the two sheets throughout the ultrasonic beam width, splits the initial node into two leaves: one for good welds and stick welds and the other for no welds and undersize welds; (ii) the relative height of the last considered echo ($h_6\_h_1$), which assesses the attenuation of the ultrasonic beam by considering the longest period of time (i.e. the time it takes to record the sixth echo), splits the good weld and stick weld node that follows the initial node into two leaves, one for good welds and the other for stick welds; (iii) the relative height of the second echo ($h_2\_h_1$) splits the no weld and undersize weld node that follows the initial node into two leaves, one for no welds and the other for undersize welds, since the second echo of an undersize weld is a *one-layer* echo whose height is much smaller than that of a no weld.

Results obtained using LOOCV cross-validation are summarized in Table 4. The classification rate of all techniques is similar, although better in the case of pruned trees, and higher than 93% for all of them. Note that the majority of misclassifications occurred between good weld and stick weld. This result is consistent with the inherent difficulty, mentioned in Martín et al. (2007), in differentiating border cases between good weld and stick weld.

Table 4. Confusion tables and classification rates of LOOCV results of CART trees using maximum and pruned trees and Gini index and Information gain splitting rules. Human operator's criterion is denoted as "original" while tree classification is denoted as "predicted".

Maximum tree - Gini index

		Predicted			
		Good	NoWeld	Stick	Undersize
Original	Good	208	0	12	1
	NoWeld	0	29	0	1
	Stick	10	1	53	1
	Undersize	0	0	2	120
	Classification rate	0.9361			

Maximum tree - Information gain

		Predicted			
		Good	NoWeld	Stick	Undersize
Original	Good	211	0	9	1
	NoWeld	0	29	0	1
	Stick	8	1	55	1
	Undersize	2	0	3	117
	Classification rate	0.9406			

Pruned tree - Gini index

		Predicted			
		Good	NoWeld	Stick	Undersize
Original	Good	213	0	7	1
	NoWeld	0	29	0	1
	Stick	9	1	54	1
	Undersize	0	0	2	120
	Classification rate	0.9498			

Pruned tree - Information gain

		Predicted			
		Good	NoWeld	Stick	Undersize
Original	Good	212	0	8	1
	NoWeld	0	29	0	1
	Stick	8	1	55	1
	Undersize	2	0	3	117
	Classification rate	0.9429			

4.2 Random forests

Random forests were implemented with the "randomForest" R package (Liaw and Wiener, 2002). In Fig. 11 OOB error and LOOCV cross-validation are presented in order to make the results comparable with the CART trees. Although results in both cases are comparable, the computational effort to obtain OOB error is much lower. The influence of the parameters *ntree* (number of trees in the forest) and *mtry* (number of features used in each node, out of the M possible features) were analysed.

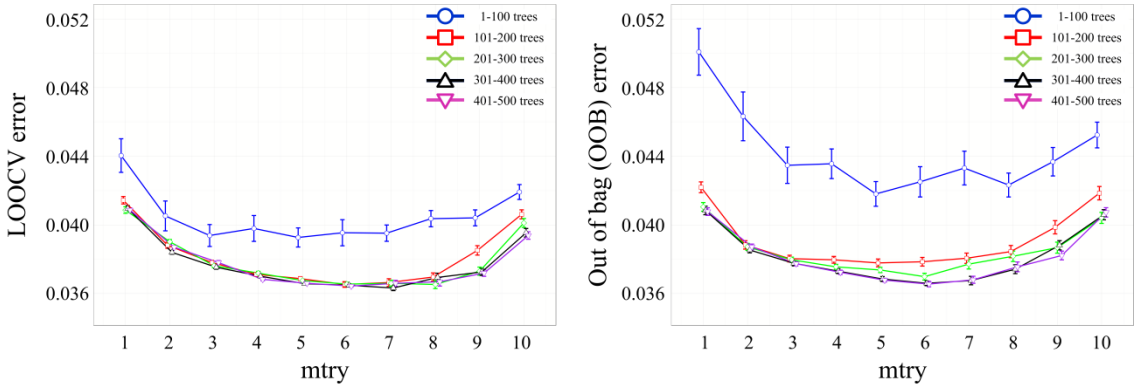

Fig. 11. Effect of *mtry* in LOOCV misclassification error (left) and in OOB misclassification error (right) depending on the number of the trees of the forest. Vertical bars show the standard errors.

The experiments showed that the higher the number of trees in the forest, the better the results obtained; however, results did not improve significantly beyond 300 trees. The best results were obtained for *mtry* equal to 6. This value corresponds approximately to $2\sqrt{M}$. In this case OOB error was about 0.03653.

All the trained random forests with more than 100 trees produced better results than any CART tree. Table 5 shows the confusion table of results of a random forest with *mtry*=6 and *ntree*=500. The classification rate was increased in about 2% in comparison to CART models, reaching an impressive 96%. Note that the most frequent misclassification was the classification of stick welds as good welds.

Table 5. Confusion table and classification rate of LOOCV results for a random forest with *mtry*=6 and *ntree*=500. Human operator's criterion is denoted as "original", while random forest classification is denoted as "predicted".

Random Forest *mtry*=6 and *ntree*=500					
		Predicted			
		Good	NoWeld	Stick	Undersize
Original	Good	218	0	3	0
	NoWeld	0	29	0	1
	Stick	6	1	57	1
	Undersize	4	0	0	118
	Classification rate		0.96347		

The disaggregated misclassification rate using the OOB approach of a random forest with *mtry*=6 was also analysed. The results are summarized in Fig. 12. In this case, errors were stabilized for less than 200 trees, and it can be observed that a considerable source of error

comes from the misclassification of stick welds. These results are consistent with those shown in Table 5.

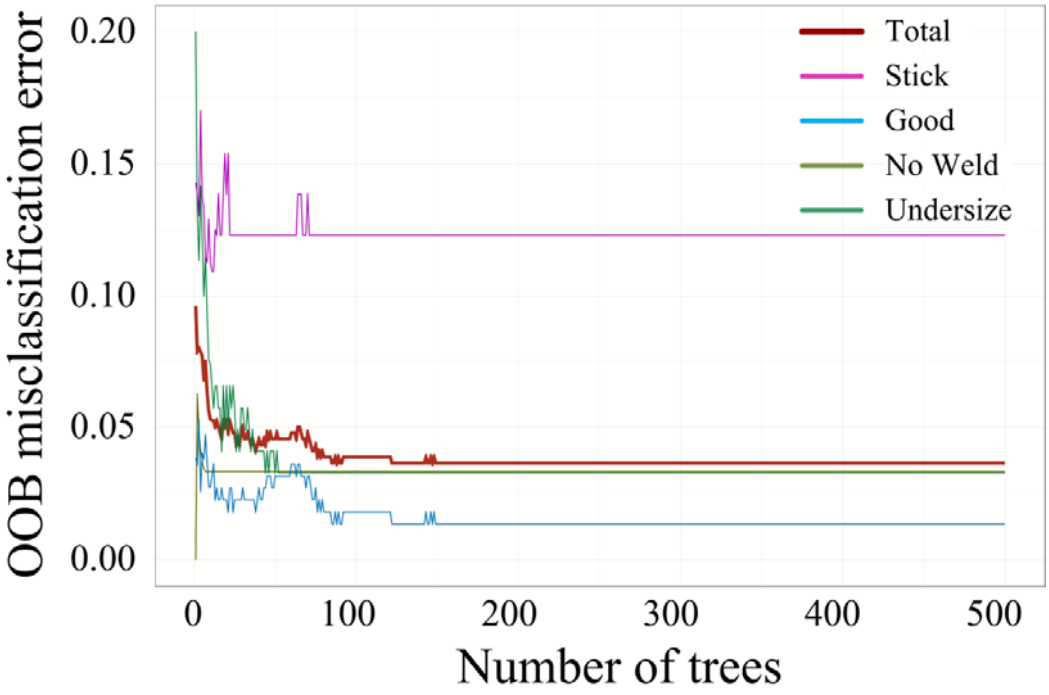

Fig. 12. OOB misclassification error of random forests with *mtry*= 6 and different number of trees depending on the quality level of the RSW joint.

A measure of variable importance in a random forest with *mtry*=6 and *ntree*=500 is shown in Table 6. The average difference over all trees between the error rate on the OOB portion of the data and the error rate after permuting each predictor variable, normalized by the standard deviation of the differences, was calculated. These results showed that the relative heights of the second ($h_2\_h_1$) and last ($h_6\_h_1$) considered echoes, together with the distance between the second and third echoes (d_2) had a key influence in the classification. This is consistent with the classification scheme followed by the decision trees shown in Figs. 9 and 10, where these three features were the first ones used in each decision tree.

Table 6. Variable importance of the features for classification accuracy.

Mean Decrease Accuracy	Error
$h_2\_h_1$	50.24
$h_6\_h_1$	42.92
d_2	35.10
$h_5\_h_1$	31.98
$h_4\_h_1$	24.68
d_3	24.61
$h_3\_h_1$	20.75
d_1	18.83
d_4	15.64
d_5	13.12

The degree of agreement, i.e. the consensus obtained by the random forest in the classification of a datum, can be used to identify the most difficult situations in which the interpretation of the ultrasonic oscillogram is least clear. This hypothesis was tested comparing the classification rate as a function of the agreement of the most voted option. Fig. 13 shows the results: 85% of the data had a consensus above 0.85 of agreement, with a classification rate of 99% of correct classification. Lower levels of agreement decreased the overall classification rate. An increase of the needed agreement of the most voted option decreased the percentage of classified joints since the criterion to take a decision was more demanding; as a consequence, the subset of classified joints had a higher classification rate.

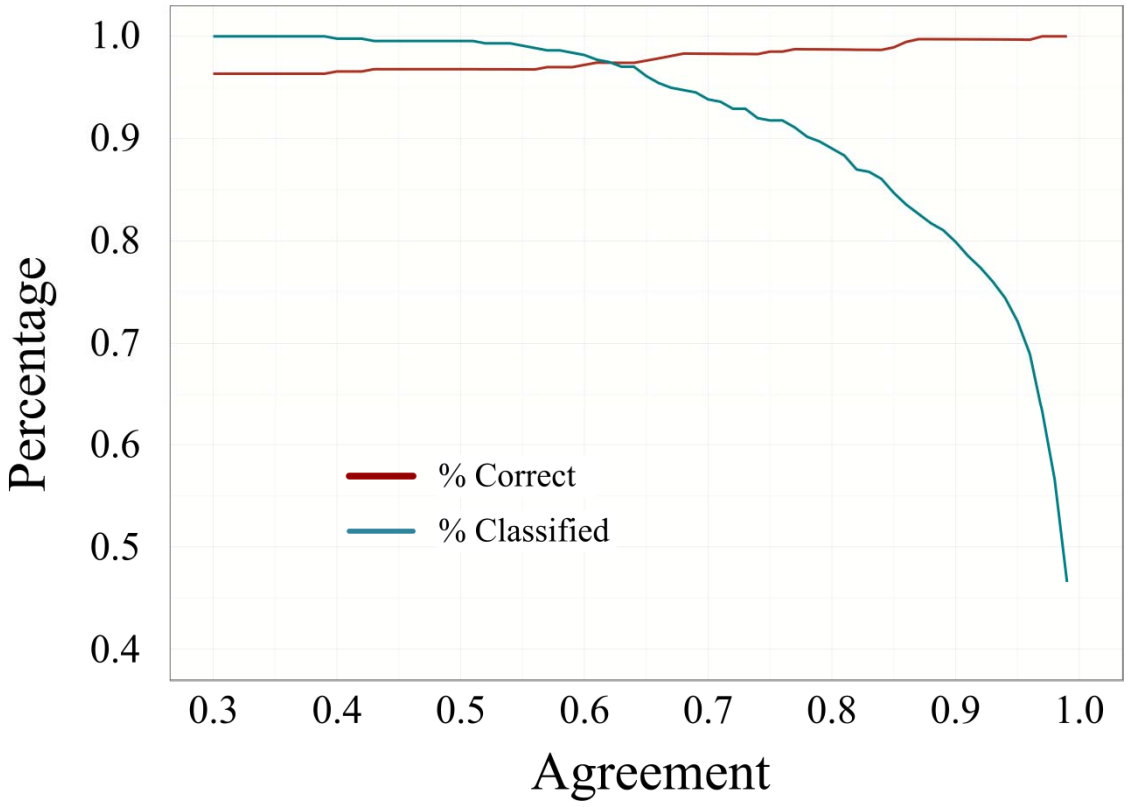

Fig. 13. Percentages of RSW joints classified and correctly classified depending on the level of agreement of the trees that constitute the random forest.

5 Conclusions

This work proposed CART and random forest techniques as pattern recognition tools for classification of ultrasonic oscillograms of RSW joints. The major conclusions are:

(1) Tree-based approaches provide good classification performance and can be employed as effective decision support tools to assist in quality control.

(2) CART trees bring to light a formal decision model compatible with the implicit classification algorithm applied by the human operator. Their interpretability, make trees very advantageous to understand and control the decision processes, instruct other human operators, compare margins of safety used by different classifiers, and establish new ones depending on the criticality of the industrial process.

(3) A random forest can aggregate decision mechanisms of several human operators, reduce overfitting and improve significantly the classification performance, although at the cost of losing interpretability and understanding of the criteria used by the decision algorithm.

(4) The democratic nature of the algorithm was applied to assess the certainty about the classification. This feature is beneficial in control quality since it reduces the workload of human operators by focusing only on controversial data which are hard to interpret.

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

Aslanlar, S., 2006. The effect of nucleus size on mechanical properties in electrical resistance spot welding of sheets used in automotive industry. Materials & Design 27, 125-131. Doi: 10.1016/j.matdes.2004.09.025

Barrera Cardiel, G., Fabián Alvarez, M.A., Vélez Martínez, M., Villaseñor, L., 2001. Inteligencias artificiales y ensayos ultrasónicos para la detección de defectos. Revista de Metalurgia 37, 403-411. Doi: 10.3989/revmetalm.2001.v37.i3.506

Breiman, L., 1996. Bagging predictors. Machine Learning 24, 123-140. Doi: 10.1007/BF00058655

Breiman, L., 2001. Random Forests. Machine Learning 45, 5-32. Doi: 10.1023/A:1010933404324

Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and Regression Trees. Chapman & Hall-CRC, Boca Raton, FL.

Chen, Z., Shi, Y., Jiao, B., Zhao, H., 2009. Ultrasonic nondestructive evaluation of spot welds for zinc-coated high strength steel sheet based on wavelet packet analysis. Journal of Materials Processing Technology 209, 2329-2337. Doi: 10.1016/j.jmatprotec.2008.05.030

Criminisi, A., Shotton, J., Konukoglu, E., 2011. Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning. Foundations and Trends in Computer Graphics and Computer Vision 7, 81-227. Doi: 10.1561/0600000035

De'Ath, G. Fabricius, K.E., 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. Ecology 81, 3178-3192. Doi: 10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2

Georgilakis, P.S., Gioulekas, A.T., Souflaris, A.T., 2007. A decision tree method for the selection of winding material in power transformers. Journal of Materials Processing Technology 181, 281-285. Doi: j.jmatprotec.2006.03.036

Hamidinejad, S.M., Kolahan, F., Kokabi, A.H., 2012. The modeling and process analysis of resistance spot welding on galvanized steel sheets used in car body manufacturing. Materials & Design 34, 759-767. Doi: 10.1016/j.matdes.2011.06.064

Hastie, T., Tibshirani, R., Friedman, J.H., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. Springer, New York, NY.

ISO 14273:2000, 2000. Specimen dimensions and procedure for shear testing resistance spot, seam and embossed projection welds.

Kaminski, R., 1997. Ultrasonic testing of spot-welded joints on coated steel sheets and optimization of welding parameters. Krautkrämer GmbH document SD 296. Accessed at http://www.spotweldtesting.com/schweisspunktpruefung.de/english/sd_296_en.pdf.

Krautkrämer, J. Krautkrämer, H., 1990. Welded Joints. In: Krautkrämer, J. and Krautkrämer, H. (Eds.), Ultrasonic Testing of Materials, 4th ed. Springer-Verlag, Berlin, pp. 431-465.

Lancaster, J.F., 1999. Metallurgy of Welding 6th ed. Abington Publishing, Cambridge, England. 229-233.

Liaw, A. Wiener, M., 2002. Classification and Regression by randomForest. R News 2, 118-122.

Mansour, T., 1991. Ultrasonic testing of spot welds in thin gage steel. In: McIntire, P. (Ed.), Nondestructive Testing Handbook. Vol. 7: Ultrasonic Testing, 2nd ed. American Society for Nondestructive Testing, Metals Park, OH, pp. 557-568.

Martín, O., 2004. Neural networks used for the optimization of quality control processes by ultrasonics in resistance spot welding joints (in Spanish). University of Valladolid (PhD Thesis),

Martín, O., López, M., De Tiedra, P., Juan, M.S., 2008. Prediction of magnetic interference from resistance spot welding processes on implantable cardioverter-defibrillators. Journal of Materials Processing Technology 206, 256-262. Doi: j.jmatprotec.2007.12.021

Martín, O., López, M., Martín, F., 2006. Redes neuronales artificiales para la predicción de la calidad en soldadura por resistencia por puntos. Revista de Metalurgia 42, 345-353. Doi: 10.3989/revmetalm.2006.v42.i5.32

Martín, O., López, M., Martín, F., 2007. Artificial neural networks for quality control by ultrasonic testing in resistance spot welding. Journal of Materials Processing Technology 183, 226-233. Doi: 10.1016/j.jmatprotec.2006.10.011

McCauley, R.B., Bennett, M.P., Bodary, W.D., Farrington, G.C., Gasser, R.J., Hurd, W.W., Schueler, A.W., Shearer, T.W., Silverberg, J.B., 1971. Resistance Spot Welding. in: Lyman, T. (Ed.), Metals Handbook 8th ed., vol. 6, Welding and Brazing American Society for Metals, Metals Park, OH, pp. 401-424.

Mitchell, M.W., 2011. Bias of the Random Forest Out-of-Bag (OOB) Error for Certain Input Parameters. Open Journal of Statistics 1, 205-211. Doi: 10.4236/ojs.2011.13024

Moshayedi, H. Sattari-Far, I., 2012. Numerical and experimental study of nugget size growth in resistance spot welding of austenitic stainless steels. Journal of Materials Processing Technology 212, 347-354. Doi: 10.1016/j.jmatprotec.2011.09.004

Özyürek, D., 2008. An effect of weld current and weld atmosphere on the resistance spot weldability of 304L austenitic stainless steel. Materials & Design 29, 597-603. Doi: 10.1016/j.matdes.2007.03.008

Quinlan, J.R., 1993. C4.5. Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, CA.

Raoelison, R., Fuentes, A., Rogeon, P., Carré, P., Loulou, T., Carron, D., Dechalotte, F., 2012. Contact conditions on nugget development during resistance spot welding of Zn coated steel sheets using rounded tip electrodes. Journal of Materials Processing Technology 212, 1663-1669. Doi: 10.1016/j.jmatprotec.2012.03.009

Rokach, L. Maimon, O.Z., 2008. Data Mining with Decision Trees: Theory and Applications. World Scientific, New Jersey.

Rupin, F., Blatman, G., Lacaze, S., Fouquet, T., Chassignole, B., 2014. Probabilistic approaches to compute uncertainty intervals and sensitivity factors of ultrasonic simulations of a weld inspection. Ultrasonics 54, 1037-1046. Doi: 10.1016/j.ultras.2013.12.006

Spinella, D.J., Brockenbrough, J.R., Fridy, J.M., 2005. Trends in Aluminum Resistance Spot Welding for the Auto Industry. Welding Journal 84, 34-40.

Therneau, T.M. Atkinson, E.J., 1997. An introduction to recursive partitioning using the rpart routines. Division of Biostatistics 61 Mayo Foundation,

Thornton, M., Han, L., Shergold, M., 2012. Progress in NDT of resistance spot welding of aluminium using ultrasonic C-scan. NDT & E International 48, 30-38.

Zhang, G. Li, G., 2012. Novel Multiple Markers to Distinguish Melanoma from Dysplastic Nevi. PLoS ONE 7, e45037. Doi: 10.1371/journal.pone.0045037

Zhang, H., 2011. Evaluation and Quality Control of Resistance-Welded Joints. in: Lienert, T., Siewert, T., Babu, S., and Acoff, V. (Eds.), Welding Fundamentals and Processes, ASM Handbook, vol. 6A ASM International, Materials Park, OH, pp. 486-504.