

Quality Prediction of Resistance Spot Welding Joints From Pre- and Post-Welding Inputs Using Stacking of Tree Ensemble Models

Óscar Martín (✉ oml@eii.uva.es)

Universidad de Valladolid <https://orcid.org/0000-0001-9200-5797>

Virginia Ahedo

Universidad de Burgos

José Ignacio Santos

Universidad de Burgos

José Manuel Galán

Universidad de Burgos

Research Article

Keywords: Resistance spot welding, Quality control, Welding parameters, Ultrasonic testing, Tree ensembles, Stacking.

Posted Date: July 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-645372/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Quality prediction of resistance spot welding joints from pre- and post-welding inputs using stacking of tree ensemble models

Óscar Martín^{a,*}, Virginia Ahedo^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. 47011, Valladolid, Spain.

<http://orcid.org/0000-0001-9200-5797>

^b Departamento de Ingeniería de Organización. Universidad de Burgos. Escuela Politécnica Superior. Edificio A. Avenida Cantabria s/n. 09006, Burgos, Spain.

Virginia Ahedo <https://orcid.org/0000-0002-9812-388X>

José Ignacio Santos <https://orcid.org/0000-0002-6653-043X>

José Manuel Galán <https://orcid.org/0000-0003-3360-7602>

* Corresponding author. Ph.: +34-983185037, , E-mail: oml@eii.uva.es (Ó. Martín)

Quality prediction of resistance spot welding joints from pre- and post-welding inputs using stacking of tree ensemble models

Abstract

Resistance spot welding (RSW) is a widespread manufacturing process in the automotive industry. There are different approaches for assessing the quality level of RSW joints. Multi-input-single-output methods, which take as inputs either the intrinsic parameters of the welding process or ultrasonic nondestructive testing variables, are commonly used. This work demonstrates that the combined use of both types of inputs can significantly improve the already competitive approach based exclusively on ultrasonic analyses. The use of stacking of tree ensemble models as classifiers dominates the classification results in terms of accuracy, F-measure and area under the receiver operating characteristic curve metrics. Through variable importance analyses, the results show that although the welding process parameters are less relevant than the ultrasonic testing variables, some of the former provide marginal information not fully captured by the latter.

Keywords: Resistance spot welding; Quality control; Welding parameters; Ultrasonic testing; Tree ensembles; Stacking.

Introduction

Manufacturers in the automotive industry face an increasingly competitive environment [1]. Remarkably, resistance spot welding (RSW) is a critical manufacturing technology in such sector [2,3], since its high speed and adaptability for automation render it suitable for mass production [4]. The number of RSW joints per vehicle is very high (around 5000 according to Xia et al. [5]), and there can be significant variability in the quality of each of them due to the fact that RSW is a complex process [6,7]; more precisely, the heat generated by an electrical current has to be substantial enough to promote local melting and the formation of a weld nugget at the faying interface [8,9], while at the same time the amount of heat is influenced by the electrical conductivity of the materials to be joined and by their surface condition, as well as by the thermal conductivity of the electrodes that are water-cooled and act as heat sinks [9].

RSW holds a promising optimization potential as a result of the balance that it establishes between cost and performance; remarkably, the tendency in the automotive industry is to reduce the number of RSW joints per vehicle, which makes the accuracy of the tools to assist in the quality control of RSW joints [10] more critical, as the fewer the RSW joints per vehicle, the stronger the requirements for each of them [6].

In the literature, two main RSW quality-control modelling approaches are found: (i) models that predict the quality of the RSW joints from the welding parameters established prior to or during the welding process [11–16], and (ii) models that assess the quality level based on the data of the ultrasonic oscillograms obtained after the welding process [6,17,18]. Notably, even though ultrasonic nondestructive testing requires that human operators have a certain degree of training, it is a promising technique for estimating the quality of RSW joints in the automotive industry [19], where it represents a cost-reduction opportunity [20,21]. As a result of the competitive nature of such sector, optimization and cost reduction are two of its cornerstones,

being thus of interest to thoroughly explore the potential and limitations of the different machine learning techniques to assess the quality level of RSW joints from either pre-welding inputs (welding parameters), post-welding inputs (ultrasonic oscillograms) or both, as well as the possible synergies between both types of information.

In the present contribution, different classification algorithms are explored and compared. Remarkably, some of the techniques that combine several classifiers simultaneously and that are showing a major success in a wide range of scientific fields have been included in the analyses. The comparison is performed using different metrics. The accuracy (i.e., the classification rate) of the algorithms is analyzed, but also the F-Score or F-measure –i.e., the harmonic mean of the precision and recall– that combines both the positive predictive value and the sensitivity in a single measure. In addition, since in different industrial environments type I error (false positives) and type II error (false negatives) do not have the same impact in terms of safety and monetary costs, the analysis was completed with the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) measure for non-binary classifiers. This latter metric provides an overall comparison of the performance of each classifier over the full range of trade-offs between the two types of errors –and hence between safety and economic costs–.

2. Experimental procedure

2.1. Materials and equipment

The chemical composition and mechanical properties of the steel sheets welded by RSW are shown in Table 1 and Table 2 respectively. Sheet thickness was 1 mm.

Table 1. Chemical composition of the steel sheet (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 steel sheet.

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 means of water-cooled truncated cone electrodes with 16 mm body diameter and 5 mm face diameter.

The ultrasonic testing of the RSW joints employed a transducer whose frequency and diameter were, respectively, 20 MHz and 4.5 mm.

2.2 Welding parameters

In this study, a total of 437 joints were obtained by RSW. The welding parameters considered include: (i) the welding current (with values varying between 4 and 8 kA RMS); (ii) the welding time (with values ranging between 4 and 20 cycles); (iii) the electrode force (whose value was kept fixed at 980.7 N); (iv) the electrode material (two types [22]: Class 2 and Class 3 of RWMA Group A); and (v) the treatment applied to the electrode material (with three options [23–25]: O61, TH02 and TF00).

2.3 Ultrasonic testing and quality levels

Ultrasonic testing was performed in accordance with the pulse-echo method with A-scan technique and, therefore, the obtained ultrasonic oscillogram is a plot of wave amplitude versus time [26]. The location of the reflecting interface determines the echo positions, and the sound attenuation –which depends on the weld nugget microstructure– determines the height of the echoes [6].

The quality level of each RSW joint was assessed from its ultrasonic oscillogram that, in turn, depends on the weld nugget; recall that the weld nugget has melted and solidified and, thus, it has a cast microstructure with coarse and columnar grains that produces higher attenuation on the ultrasonic beam than that of the parent metal. Thereupon, the thickness of the weld nugget influences the ultrasonic oscillogram in the following manner: the greater the thickness, the higher the attenuation. Another parameter of the weld nugget, its diameter, affects the ultrasonic oscillogram as well; more precisely, given that the interface between the two steel sheets causes the reflection of the ultrasonic beam, one-layer echoes will appear between principal echoes if the weld nugget diameter is smaller than the ultrasonic beam width [27].

Considering the effect of weld nugget on the ultrasonic beam behaviour, four quality levels were established for classifying RSW joints [6,28]:

- **Good weld.** The thickness of the weld nugget is large and its diameter is greater than the ultrasonic beam width. Therefore, the span of the sequence of echoes is short due to the high attenuation, and the distance between consecutive echoes is the sum of the thickness of each of the two steel sheets.
- **Undersize weld.** The diameter of the weld nugget is smaller than the ultrasonic beam width. Thus, the portion of the ultrasonic beam that does not pass through the weld nugget –and whose reflection takes place at the interface between the two steel sheets– causes one-layer echoes between the principal echoes.
- **Stick weld.** The weld nugget has an adequate diameter but a small thickness. Hence, the distance between consecutive echoes is the sum of the thickness of each of the two steel sheets, and the attenuation is lower than in a good weld, so the span of the sequence of echoes is longer.
- **No weld.** There is no melted and solidified metal. Therefore, the span of the sequence of echoes is longer than that of a RSW joint with weld nugget, and the distance between echoes is the thickness of one steel sheet.

2.4 Inputs for assessing the quality level

A total of 14 inputs were considered for assessing the quality level (Fig. 1):

- Four of the five welding parameters: welding current, welding time, electrode material, and treatment applied to the electrode material (the electrode force was not considered as an input because its value was kept fixed for all RSW joints).
- The ten components of the representative vector of each ultrasonic oscillogram, obtained with a program developed by Martín [29]. This 10-component vector uses only the first six echoes of the ultrasonic oscillogram [6,17]:
 - The first five components of the vector are the relative heights of the echoes. Specifically, the n th component (with $n=1, \dots, 5$) is the height of the $[n + 1]$ th echo (h_{n+1}) divided by the height of the 1st echo (h_1).
 - The last five components of the vector are the distances between consecutive echoes: the n th component (with $n=6, \dots, 10$) is the distance between the $[n-4]$ th echo and the $[n-5]$ th echo (d_{n-5}).

2.5 Computational methods

The experimental design used to assess the influence of the individual and combined use of pre-welding and post-welding inputs as well as the effectiveness of the different classification tools and algorithms in exploiting such information is as follows. The performance of a large number of classifiers—including tree ensembles and stacking models—has been analyzed using stratified nested cross-validation (10 folds). The hyperparameters of the algorithms have been optimized through grid search in the inner loop, and their performance has been measured in the outer loop. In addition, variable importance analyses based on a random forest model were conducted to determine the relative impact of each of the inputs in the overall classification accuracy attained.

A brief description of the main classifiers and of the different variable importance analyses conducted is provided below.

2.5.1 AdaBoost

Due to its great accuracy on a multitude of very diverse problems [30,31], boosting is one of the most popular classification techniques. In this paper, two of the most prominent boosting algorithms are used: Adaptive Boosting (AdaBoost) [32]—the most popular and well-known boosting algorithm—and the eXtreme Gradient Boosting (XGBoost) algorithm [33], which, as a consequence of its exceptional results, has received a lot of attention since its publication.

Boosting consists in obtaining a strong classifier from the sequential combination of weak base learners. AdaBoost typically uses classification trees as base classifiers (in the present work, two types of base classifiers were used: decision stump and J48 trees) and has been employed in identifying and classifying weld defects [34]. At each iteration, the training set is reweighted so that higher weights are given to instances or data that have been previously misclassified, remaining the previous classifiers unchanged. This distribution of weights is denoted by D_t and changes at each iteration t of the algorithm. Each time AdaBoost trains a tree on the training sample $\{(x_i, y_i)\}$, where x_i are the regressors and y the output variable, it generates a weak

hypothesis h_t aimed at minimizing the error of the distribution D_t through the weak classifier Eq. (1).

$$\varepsilon_t = \Pr_{i \sim D_t}[h_t(x_i) \neq y_i] \quad (1)$$

AdaBoost then estimates an α_t parameter according to Eq. (2), which weights the contribution of classifier t to the overall strong classifier.

$$\alpha_t = \log\left(\frac{1 - \varepsilon_t}{\varepsilon_t}\right) \quad (2)$$

The next step is updating the weight distribution so as to train the next weak classifier. In the case of multi-class classification, if the M.1 version of the algorithm [32] is used, this weighting is performed according to Eq. (3), where \mathbf{I} is the indicator function.

$$D_{t+1}(i) = D_t(i) \exp(\alpha_t \mathbf{I}(h_t(x_i) \neq y_i)) \quad (3)$$

When the algorithm has built the number of trees selected, it generates the final classification using Eq. (4), where T is the total number of trained trees.

$$H(x) = \text{sign}\left(\sum_{t=1}^T \alpha_t h_t(x)\right) \quad (4)$$

2.5.2 XGBoost

XGBoost is an algorithm that also uses boosting, but in this case, based on the additive expansion of regression trees. More specifically, it uses gradient-boosting decision trees as a base, but incorporates different mechanisms to exploit memory resources and parallelization, thus reducing computation times efficiently [33]. XGBoost gained fame by winning 17 of the 29 machine learning tasks proposed in Kaggle in 2015, a fact that together with its current high cross-platform portability, has resulted into XGBoost receiving a lot of attention. Although it is an algorithm that reaches its maximum potential in massive datasets –as it makes a more efficient use of computational resources than other algorithms– its application to small and medium-sized datasets is also beginning to receive attention, e.g. in welding processes [35]. From a formal perspective, and given a sample of n data and m features, where again x_i are the regressors and y the output variable, the objective function of the algorithm is:

$$\mathcal{L} = \sum_i^n l(y_i, \hat{y}_i) + \sum_{t=1}^T \Omega(f_t) \quad (5)$$

Where l is a differentiable convex loss function that captures the difference between the prediction and the actual data, f_t is the t -th tree in the additive expansion and $\Omega(f_t)$ is a regularization function that penalizes the complexity of the regression trees. The loss function is approximated through the second-order Taylor expansion:

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^n \left[l(y_i, \hat{y}^{(t-1)}) + g_i f_i(x_i) + \frac{1}{2} h_i f_i^2(x_i) \right] + \Omega(f_t) \quad (6)$$

Where g_i represents the first derivative of each sample (gradient) and h_i indicates the second derivative of each sample (hessian). In the specific case of multi-class classification used in this work, the loss function is a generalization of the logistic loss function [30].

2.5.3 Random Forest

The random forest algorithm is based on bootstrap aggregation (bagging), that is, it combines the results obtained by multiple classification trees built –deep and unpruned– on different bootstrapped samples as a way to reduce variance and hence improve accuracy. However, it outperforms bagging by considering just a subset of predictors (random subspace method) at each split, which serves to decorrelate the different trees in the forest [36,37]. In classification problems, the number of predictors (m) typically considered at each split is $m \approx \sqrt{p}$, where p is the total number of predictors. Notably, random forest models are robust to overfitting and to the presence of correlated regressors. In addition, even though bootstrap aggregation methods result in improved accuracy at the expense of interpretability, they enable the obtention of variable importance measures such as the original individual variable importance proposed by Breiman [37], group variable importance [38] and conditional variable importance [39,40] and, thus, they have been successfully applied to quality assessment in resistance spot welding process [6,41].

2.5.4 Individual variable importance

Within the framework of random forests for classification, individual variable importance analysis is generally enacted as follows: once the model is built, each bootstrapped sample has a tree fitted to it and the corresponding out-of-bag (OOB) sample –approximately one third of the observations in the real dataset [36]–. Let us assume that such real dataset has M predictors; to determine the relative importance of the m th predictor, its values are randomly permuted in all the OOB samples and then run down their corresponding trees. As a result, each OOB observation obtains several class label predictions –the quantity of which depends on the number of OOB samples where it appears–; eventually, the majority vote is taken and it is compared with the true class label to compute the misclassification rate. The importance of the m th predictor is subsequently calculated as the change in classification accuracy after the permutation with respect to the original case –mean decrease in accuracy over all trees [37]–.

2.5.5 Conditional variable importance

In [39], the authors pointed that the above-described variable importance measure showed a bias towards correlated predictors and developed an alternative measure: the conditional variable importance, in which the dependence between a predictor and the outcome is calculated conditionally upon the values of other predictors. In particular, for each tree, they propose to divide –completely bisect– the predictor space into a multidimensional grid in accordance with the partition induced by that tree, and it is within each such partition that the OOB values are conditionally permuted. Eventually, the importance of each variable is calculated as the difference in predictive accuracy before and after the permutation, and averaged across all trees [40] –as in Breiman’s measure–.

In the present case study, conditional variable importance is of interest since it is particularly appropriate to address the research question of whether an increase in predictive accuracy may be attained as a consequence of considering both the welding parameters and the oscillogram variables together.

2.5.6 Group variable importance

Sometimes, rather than the individual importance of each variable, it is the joint importance of different groups of variables that is of interest. For such cases, Gregorutti et al. [38] proposed a group variable importance method that consists of using the same random permutation for each variable in the group under consideration, so that the empirical joint distribution of the group of variables is preserved, but the link between such group of variables, the rest of predictors and the response is effectively broken. In the present contribution, two groups of variables were considered: (i) welding parameters and (ii) oscillogram variables.

2.5.7 Rotation Forest

Rotation forest is a tree-based ensemble method that seeks both diversity and accuracy by means of feature extraction. More specifically, rotation forest uses the C4.5 decision tree [42] as base classifier and it is defined by three main parameters: (i) L –the number of trees in the forest–; (ii) K –the number of subsets in which the feature set is split–; and (iii) p –the proportion of the observations (X) to select–. The training set for each classifier is built as follows: first, the feature space is divided into K –disjoint– subsets; then, for every such subset a nonempty subset of classes is randomly selected and a bootstrap sample of size 75% of the data count is drawn; subsequently, Principal Component Analysis (PCA) is run on the reduced dataset that includes only the features in the corresponding subset and the bootstrapped observations selected; eventually, the PCA coefficients obtained across all subsets are rearranged in a rotation matrix (R_i^a) so as to match the original feature order, and the classifier is built using (XR_i^a, Y) as the training set [43].

Remarkably, for problems with continuous real-valued features, rotation forest was found to be significantly more accurate on average than competing techniques from the families of algorithms: support vector machines (SVMs), tree-based ensembles and neural networks [44], being thus recommended to consider it among the algorithms with the greatest performance.

2.5.8 Other classifiers

The rest of classifiers implemented can be listed and succinctly described as follows: (i) Naive Bayes is a probabilistic classifier based on Bayes' theorem that, despite being a simple Bayesian model and making the assumption that given the class each feature is independent of any other, has proven to give good results in many real contexts [45]. (ii) Several implementations and variants of SVM classifiers, which maximize the width of the gap between different classes by mapping the data into a higher-dimensional space where the separation of the classes is simpler –recall that for the mapping process, it is possible to use different kernels–; more specifically, in the present contribution we used the Least Squares Support Vector Machine classifier (LSVM) [46] with radial basis function (RBF) kernel, the sequential minimal optimization algorithm for

training a support vector classifier [47,48] with radial basis function (RBF) and the Pearson VII function-based universal kernel (PUK) [49]. (iii) The performance of a multilayer perceptron neural network [50], (iv) logistic regression [51], and (v) the J48 algorithm –a Java implementation of the C4.5 decision tree [42]– was also analyzed. Eventually, commonly used baselines based on simple decision rules such as (vi) OneR and (vii) ZeroR were also evaluated, to compare against them the behavior of the more sophisticated classifiers [51] implemented.

2.5.9 Stacking

Apart from bagging and boosting [52] –the two most common ensemble techniques for classification– there is a third methodology for combining classifiers on the same dataset known as stacked generalization or stacking [53]. This approach is giving excellent and, in some cases, close to optimal results [54–56], and it is also recently starting to be successfully applied in manufacturing-related fields [57,58]. While the base classifiers are usually of the same type as in the other ensemble techniques, the idea of stacking is the opposite. Typically, stacking consists of combining the strengths of classifiers of different nature and based on dissimilar hypotheses to generate a more accurate classifier. A two-level structure is then used to decide in which cases to use the predictions of each algorithm. Initially, different base classifiers are trained –the level-0 models–. The predictions of these models constitute the input of another classifier known as the metalearner or level-1 model, which learns a good generalization of the classifiers it combines. The primary goal of the metalearner is to detect the regions of the classification space in which each classifier or set of classifiers is most reliable. Although, in principle, any level-1 generalizer that is relatively global and smooth can be expected to perform well, in practice it is often appropriate to use overfitting-resistant classifiers such as logistic regression or random forests –which are the ones used in this work–. In the implementation selected, the metalearner receives as attributes the vector with the probabilities of each class, a strategy that increases the generalization capacity [59].

Since to be able to generalize the results of the metalearner it is necessary not to use the same instances as at level 0, nested cross-validation was conducted; more specifically, 5-fold internal cross-validation was implemented within the inner loop, being the honest evaluation of the classifier performance conducted on the outer loop.

3 Results

For each performance metric (accuracy, F-measure and AUC) an ANOVA test was conducted to check the null hypothesis of equality of means across the 50 algorithms implemented. In accordance with the results obtained, the null hypothesis can be rejected in all cases at a significance level of 0.001, which means that some algorithms perform significantly better than the baseline classifiers.

A post hoc analysis was then performed for each metric using Duncan's multiple range test at a 0.1 level of significance (a corrected paired Student's *t*-test [60] was also performed and it gave very similar results at 0.05 significance level). Recall that in Duncan's multiple range test, two classifiers are considered statistically different if their difference exceeds the studentized range statistic. The results are presented in Table 3. Differences between performance metrics that do not share the same letter are considered statistically significant.

Table 3. Results of the algorithms using different input variables –welding parameters (WP) and ultrasonic oscillograms (US)– in terms of accuracy. F-measure and AUC using stratified 10-fold nested cross-validation. Rows are ordered according to accuracy. In the "Algorithm" column, in stacking algorithms, the metalearner is specified in parentheses. The standard deviation of each measure is shown in parentheses. The significance groups of the post hoc analysis Duncan multiple range test for a significance level of 0.1 are also represented in the table. Two algorithms are statistically different at that level if they do not share the same letter.

Algorithm	Dataset	Accuracy	Groups	F-measure	Groups	AUC	Groups
Stacking(RF) RF-XGBoost	US + WP	97.717 (1.87)	a	0.9822 (0.018)	a	0.9965 (0.006)	a
Random Forest	US + WP	97.257 (1.45)	ab	0.9778 (0.015)	ab	0.9983 (0.002)	a
Stacking(RF) Ensemble trees	US + WP	97.257 (1.43)	ab	0.9776 (0.015)	ab	0.9986 (0.001)	a
Adaboost(J48)	US + WP	97.252 (1.80)	ab	0.9754 (0.017)	ab	0.9767 (0.015)	abc
Stacking(Logistic) Ensemble trees	US + WP	97.030 (1.53)	ab	0.9730 (0.014)	ab	0.9992 (0.001)	a
Rotation Forest	US + WP	97.024 (1.11)	ab	0.9755 (0.013)	ab	0.9983 (0.003)	a
Stacking(RF) Ensemble trees	US	96.802 (2.44)	abc	0.9732 (0.021)	ab	0.9963 (0.008)	a
Stacking(Logistic) Ensemble trees	US	96.564 (1.62)	abcd	0.9710 (0.015)	abc	0.9981 (0.002)	a
XGBoost	US + WP	96.564 (1.94)	abcd	0.9715 (0.017)	abc	0.9979 (0.003)	a
Stacking(RF) RF-XGBoost	US	96.353 (2.19)	abcd	0.9706 (0.024)	abc	0.9944 (0.008)	ab
Random Forest	US	96.342 (2.45)	abcd	0.9711 (0.018)	abc	0.9977 (0.002)	a
Adaboost(J48)	US	96.337 (2.67)	abcd	0.9665 (0.022)	abc	0.9740 (0.020)	abc
XGBoost	US	96.337 (1.92)	abcd	0.9689 (0.016)	abc	0.9973 (0.004)	a
Rotation Forest	US	95.883 (2.79)	abcd	0.9689 (0.016)	abc	0.9979 (0.003)	a
J48	US + WP	95.661 (4.48)	abcd	0.9668 (0.022)	abc	0.9732 (0.021)	abc
J48	US	95.206 (4.47)	abcd	0.9619 (0.024)	abc	0.9731 (0.020)	abc
LSVM RBF Kernel	US	94.286 (5.05)	abcd	0.9533 (0.051)	abc	0.9519 (0.053)	c
LSVM RBF Kernel	US + WP	94.064 (2.88)	abcd	0.9666 (0.019)	abc	0.9657 (0.020)	abc
SVM RBF Kernel	US	93.832 (2.38)	bcd	0.9570 (0.020)	abc	0.9696 (0.014)	abc
Naive Bayes	US + WP	93.351 (2.78)	cd	0.9672 (0.018)	abc	0.9927 (0.009)	ab
Naive Bayes	US	93.351 (3.16)	cd	0.9715 (0.015)	abc	0.9923 (0.008)	ab
SVM Puk Kernel	US	93.145 (3.01)	d	0.9519 (0.035)	abc	0.9696 (0.020)	abc
SVM Puk Kernel	US + WP	89.255 (3.37)	e	0.9383 (0.031)	cd	0.9576 (0.020)	c
Logistic Regression	US + WP	89.017 (3.18)	e	0.9691 (0.018)	abc	0.9946 (0.006)	ab
SVM RBF Kernel	US + WP	88.784 (4.24)	ef	0.9439 (0.028)	bcd	0.9625 (0.019)	bc
Logistic Regression	US	87.178 (2.72)	efg	0.9665 (0.012)	abc	0.9918 (0.011)	ab
Multilayer Perceptron	US	85.809 (3.20)	fg	0.9198 (0.036)	d	0.9819 (0.019)	abc
Multilayer Perceptron	US + WP	85.328 (4.85)	g	0.9154 (0.029)	d	0.9724 (0.017)	abc
Rotation Forest	WP	80.090 (5.90)	h	0.8537 (0.057)	e	0.8986 (0.043)	d
XGBoost	WP	79.202 (5.23)	hi	0.8386 (0.060)	ef	0.8846 (0.043)	def
Stacking(Logistic) Ensemble trees	WP	78.959 (5.73)	hi	0.8412 (0.059)	ef	0.8803 (0.059)	defg
SVM Puk Kernel	WP	78.034 (4.82)	hi	0.8404 (0.044)	ef	0.8497 (0.051)	ghi
Stacking(RF) Ensemble trees	WP	77.585 (4.63)	hij	0.8350 (0.053)	ef	0.8886 (0.054)	de
J48	WP	77.130 (4.74)	hijk	0.8431 (0.053)	ef	0.8579 (0.055)	efgh
LSVM RBF Kernel	WP	76.903 (7.09)	hijk	0.8291 (0.067)	efg	0.8236 (0.068)	ij
AdaBoost(DS)	US + WP	76.660 (2.81)	hijk	0.8623 (0.020)	e	0.8589 (0.066)	efgh
AdaBoost(DS)	US	76.660 (2.81)	hijk	0.8623 (0.020)	e	0.8589 (0.066)	efgh
SVM RBF Kernel	WP	76.427 (4.06)	ijk	0.8358 (0.053)	ef	0.8522 (0.056)	ghi
Random Forest	WP	75.988 (4.77)	ijk	0.8134 (0.064)	fgh	0.8762 (0.048)	defgh
Adaboost(J48)	WP	74.392 (4.29)	jkl	0.7968 (0.061)	hi	0.8750 (0.051)	defgh
OneR	US + WP	74.144 (4.46)	jkl	0.8529 (0.027)	e	0.8491 (0.030)	ghi
OneR	US	74.144 (4.46)	jkl	0.8529 (0.027)	e	0.8491 (0.030)	ghi
Stacking(RF) RF-XGBoost	WP	73.689 (6.53)	kl	0.7958 (0.063)	hi	0.8439 (0.066)	hi
Multilayer Perceptron	WP	72.098 (7.67)	l	0.8031 (0.072)	ghi	0.8541 (0.083)	fghi
Naive Bayes	WP	69.123 (5.92)	m	0.7888 (0.053)	hi	0.8123 (0.050)	j
Logistic Regression	WP	67.072 (6.16)	m	0.7749 (0.054)	i	0.8103 (0.043)	j
OneR	WP	59.017 (5.11)	n	0.7227 (0.036)	j	0.6384 (0.053)	k
AdaBoost(DS)	WP	55.354 (5.32)	o	0.6645 (0.065)	k	0.6163 (0.061)	k
ZeroR	US + WP	50.803 (0.94)	p	0.6737 (0.008)	k	0.5000 (0.000)	l
ZeroR	US	50.803 (0.94)	p	0.6737 (0.008)	k	0.5000 (0.000)	l
ZeroR	WP	50.803 (0.94)	p	0.6737 (0.008)	k	0.5000 (0.000)	l

These results show several relevant aspects. First, there is a wide range of classifiers that obtain good results, i.e., whose performance is significantly superior to the 50% prediction accuracy of the ZeroR baseline. The fundamental predictive component lies in both the specific classifier selected and in the set of input variables used to train and validate the model. In this latter vein, the use of the welding parameters alone allows us to correctly classify approximately 80% of the instances using tree ensemble algorithms. In particular, Random Forest, XGBoost and Stacking of tree ensembles obtain the best results in terms of accuracy. In accordance with the AUC, these algorithms are also statistically significantly better than the rest of the algorithms on the same dataset.

Although these results are interesting per se and allow to conduct the quality control of the welding process exclusively from the welding parameters, if they are compared against a posteriori RSW joint quality analysis techniques—such as an analysis based on ultrasonic testing—, the accuracy attained in this latter case is significantly better. More precisely, if only the data from ultrasonic oscillograms is used to determine the quality of the RSW joints, a wide range of algorithms far exceed 90% accuracy; in particular, all algorithms based on trees, as well as several algorithms based on SVMs with different kernels; it should be recalled that even simpler classifiers such as Naive Bayes obtain meritorious results. Within such a framework, it is also worth highlighting that stacking algorithms based on tree ensembles already achieve extremely high prediction results.

A fundamental question addressed in the present contribution is whether the joint combination of the two RSW joint quality determination methods—i.e., the welding parameters and the ultrasonic oscillogram variables— can improve the automatic quality-level classification. For this purpose, the results provided in Fig. 2 are studied. Fig. 2 shows the accuracy results of the algorithms that obtain a 90% accuracy with any of the regressor sets used—i.e., the welding parameters alone, the oscillogram variables alone or the combination of the two—, and compares the prediction obtained in each case. A relevant result is that tree-based algorithms—even basic trees such as J48—are able to combine information from both sources—welding parameters and oscillograms—to improve the prediction by almost 1% systematically (p -value=0.0012 using paired t -test). This result is significant because improving classification when many algorithms are already performing above 95% accuracy is challenging. Other classifiers that are not as robust to partially redundant information, such as SVMs, have trouble in incorporating this information effectively, being the combination of both regressor sets even counterproductive in some cases.

An additional aspect to highlight is the excellent performance obtained by the stacking algorithms in the task of assessing the quality level of RSW joints based on either the welding parameters, the ultrasonic oscillogram variables or both. In terms of accuracy and F -measure, the combination of random forest together with XGBoost gives the best results. Still, stacking based on tree ensembles (combining random forest, rotation forest, XGBoost and adaBoost) using as metalearner either logistic regression or random forest also gives very competitive results. In the case of using the area under the curve (AUC) as the performance measure, the approach that provides the best results uses as base learners the four tree ensembles listed above with any of the two metalearners proposed. It should be recalled that the use of one or the other performance metric in industrial contexts will depend on the variation over time of the non-quality costs. Remarkably, the results obtained show that the combination of stacking with tree ensembles when both the welding parameters and the variables from ultrasonic

nondestructive testing are used as inputs, provides very accurate results for any desired range of specificity and sensitivity.

As regards the variable importance analyses conducted, Fig. 3 shows on the left the results of the individual variable importance proposed by Breiman [37] –which have been obtained using the *randomForestSRC* R package [61]– and on the right the individual conditional importance of each variable according to Debeer and Strobl [40] –which were calculated with the R packages *party* [39,62,63] and *permimp* [40]–. Notably, Breiman’s approach provides the importance of each variable within the model, as its random permutation of the values of the regressor variable is supposed to mimic the absence of such variable within the model; on its part, the conditional variable importance approach quantifies the contribution of each predictor conditioned to the presence of the rest of the regressors. Recall that even though the conditional approach induces changes in the overall ranking of all the predictors with respect to Breiman’s method, the most remarkable change is related to different position of the welding process variable *welding time*; in fact, its higher position in the conditional variable importance analysis may be interpreted in relation to the 1% increase in predictive accuracy that is attained when both the welding process and the oscillogram variables are considered.

In relation to the group variable importance analysis, its results are shown in Fig. 4. It serves to highlight that the set of most discriminant regressors is by far that of the oscillogram variables.

4 Conclusions

In this work, the use of pre-welding inputs (welding parameters) and post-welding inputs (ultrasonic oscillogram variables) to determine the quality level of RSW joints has been analyzed both individually and in combination. In the analyses, a large number of classifiers have been compared using stratified nested cross-validation, i.e., optimizing the hyperparameters of the algorithms through grid search in the inner loop, and measuring the performance in the outer loop. The results obtained show that:

- The analysis of RSW joint quality using post-welding variables is systematically superior to using only the pre-welding inputs. Compared to the baseline (ZeroR) the welding parameters increase the accuracy by an extra 30%, while the analysis based on ultrasonic testing improves the prediction by 45%. In particular, a large set of algorithms are capable of obtaining very accurate results (an accuracy greater than 90%) using ultrasonic testing data.
- The combined use of pre-welding and post-welding inputs allows improving prediction in algorithms that are less sensitive to overfitting and internal correlations between regressors. In particular, algorithms based on trees and tree ensembles allow statistically significant improvements in prediction. Given the good results of many algorithms using only ultrasonic oscillograms, improving the outcomes is particularly challenging; however, since in competitive industries such as the automotive one little improvements make the difference, and in view of the low implementation cost of tracking the welding parameters compared to using ultrasonic oscillograms exclusively, the combination of both approaches may be of interest for the reduction of RSW non-quality costs.
- The variable importance analyses conducted on the basis of random forests confirm that the information for quality evaluation of the oscillograms is superior to that of the welding parameters. However, welding parameters such as welding time and the

treatment applied to the electrode material have an influence that is still relevant and not fully captured by the ultrasonic testing, thus being of interest to exploit this information in a valuable way.

- The results show that stacking techniques that effectively combine different classifiers to issue the final prediction yield the best performance for all the prediction metrics analyzed. The joint combination of boosting- and bagging-based tree ensembles obtains the best results for the problem.

Acknowledgements

The authors acknowledge financial support from the Spanish Ministry of Science, Innovation and Universities (excellence networks HAR2017-90883-REDC and RED2018-102518-T), the Junta de Castilla y León –Consejería de Educación through BDNS 425389 and the predoctoral grant awarded to Virginia Ahedo (supported by the European Social Fund)–. The authors acknowledge support from the Santander Supercomputación group (University of Cantabria), which provided access to the Altamira Supercomputer –located at the Institute of Physics of Cantabria (IFCA-CSIC) and member of the Spanish Supercomputing Network– to perform the different simulations/analyses.

Declarations

- Funding:** Financial support from:
 - The Spanish Ministry of Science, Innovation and Universities (excellence networks HAR2017-90883-REDC and RED2018-102518-T).
 - The Junta de Castilla y León: Consejería de Educación through BDNS 425389 and the predoctoral grant awarded to Virginia Ahedo (supported by the European Social Fund).
- Conflicts of interest:** There is no conflict of interest that the authors of this work aware of.
- Availability of data and material:** The data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.
- Code availability:** Not applicable
- Authors' contributions:** All authors have contributed equally in each and every task of the article.
- Ethics approval:** Not applicable.
- Consent to participate:** Not applicable.
- Consent for publication:** Not applicable.

References

- [1] Mousavizade, S. M., and Pouranvari, M., 2019, "Projection Friction Stir Spot Welding: A Pathway to Produce Strong Keyhole-Free Welds," *Sci. Technol. Weld. Join.*, **24**(3), pp. 256–262.
- [2] Pouranvari, M., Aghajani, H., and Ghasemi, A., 2020, "Enhanced Mechanical Properties of Martensitic Stainless Steels Resistance Spot Welds Enabled by in Situ Rapid Tempering," *Sci. Technol. Weld. Join.*, **25**(2), pp. 119–126.

- [3] Ertas, A. H., and Akbulut, M., 2021, "Experimental Study on Fatigue Performance of Resistance Spot-Welded Sheet Metals," *Int. J. Adv. Manuf. Technol.*, **114**(3–4), pp. 1205–1218.
- [4] Martín, Ó., López, M., De Tiedra, P., and Juan, M. S., 2008, "Prediction of Magnetic Interference from Resistance Spot Welding Processes on Implantable Cardioverter-Defibrillators," *J. Mater. Process. Technol.*, **206**(1–3), pp. 256–262.
- [5] Xia, Y.-J., Shen, Y., Zhou, L., and Li, Y.-B., 2021, "Expulsion Intensity Monitoring and Modeling in Resistance Spot Welding Based on Electrode Displacement Signals," *J. Manuf. Sci. Eng.*, **143**(3), p. 031008.
- [6] Martín, Ó., Pereda, M., Santos, J. I., and Galán, J. M., 2014, "Assessment of Resistance Spot Welding Quality Based on Ultrasonic Testing and Tree-Based Techniques," *J. Mater. Process. Technol.*, **214**(11), pp. 2478–2487.
- [7] Bag, S., DiGiovanni, C., Han, X., and Zhou, N. Y., 2020, "A Phenomenological Model of Resistance Spot Welding on Liquid Metal Embrittlement Severity Using Dynamic Resistance Measurement," *J. Manuf. Sci. Eng.*, **142**(3), p. 031007.
- [8] Shen, Y., Xia, Y.-J., Li, H., Zhou, L., Pan, H.-T., and Li, Y., 2021, "A Novel Expulsion Control Strategy with Abnormal Condition Adaptability for Resistance Spot Welding," *J. Manuf. Sci. Eng.*
- [9] Oliveira, J. P., Ponder, K., Brizes, E., Abke, T., Edwards, P., and Ramirez, A. J., 2019, "Combining Resistance Spot Welding and Friction Element Welding for Dissimilar Joining of Aluminum to High Strength Steels," *J. Mater. Process. Technol.*, **273**(March), p. 116192.
- [10] Pereda, M., Santos, J. I., Martín, Ó., and Galán, J. M., 2015, "Direct Quality Prediction in Resistance Spot Welding Process: Sensitivity, Specificity and Predictive Accuracy Comparative Analysis," *Sci. Technol. Weld. Join.*, **20**(8), pp. 679–685.
- [11] Zhou, K., and Yao, P., 2019, "Overview of Recent Advances of Process Analysis and Quality Control in Resistance Spot Welding," *Mech. Syst. Signal Process.*, **124**, pp. 170–198.
- [12] Valaee-Tale, M., Sheikhi, M., Mazaheri, Y., Malek Ghaini, F., and Usefifar, G. R., 2020, "Criterion for Predicting Expulsion in Resistance Spot Welding of Steel Sheets," *J. Mater. Process. Technol.*, **275**(January 2019), p. 116329.
- [13] Martín, Ó., Ahedo, V., Santos, J. I., De Tiedra, P., and Galán, J. M., 2016, "Quality Assessment of Resistance Spot Welding Joints of AISI 304 Stainless Steel Based on Elastic Nets," *Mater. Sci. Eng. A*, **676**, pp. 173–181.
- [14] Martín, Ó., Tiedra, P. De, López, M., San-Juan, M., García, C., Martín, F., and Blanco, Y., 2009, "Quality Prediction of Resistance Spot Welding Joints of 304 Austenitic Stainless Steel," *Mater. Des.*, **30**(1), pp. 68–77.
- [15] Boersch, I., Füssel, U., Gresch, C., Großmann, C., and Hoffmann, B., 2018, "Data Mining in Resistance Spot Welding: A Non-Destructive Method to Predict the Welding Spot Diameter by Monitoring Process Parameters," *Int. J. Adv. Manuf. Technol.*, **99**(5–8), pp. 1085–1099.
- [16] Mikno, Z., Grzesik, B., and Stępień, M., 2020, "The Investigation on the Ideal Spot Weld Numerical Model in Resistance Welding," *Int. J. Adv. Manuf. Technol.*, **111**(3–4), pp. 895–907.

- [17] Martín, Ó., López, M., and Martín, F., 2007, "Artificial Neural Networks for Quality Control by Ultrasonic Testing in Resistance Spot Welding," *J. Mater. Process. Technol.*, **183**(2–3), pp. 226–233.
- [18] Amiri, N., Farrahi, G. H., Kashyzadeh, K. R., and Chizari, M., 2020, "Applications of Ultrasonic Testing and Machine Learning Methods to Predict the Static & Fatigue Behavior of Spot-Welded Joints," *J. Manuf. Process.*, **52**(January), pp. 26–34.
- [19] Chen, Z., Shi, Y., Jiao, B., and Zhao, H., 2009, "Ultrasonic Nondestructive Evaluation of Spot Welds for Zinc-Coated High Strength Steel Sheet Based on Wavelet Packet Analysis," *J. Mater. Process. Technol.*, **209**(5), pp. 2329–2337.
- [20] Spinella, D. J., Brockenbrough, J. R., and Fridy, J. M., 2005, "Trends in Aluminum Resistance Spot Welding for the Auto Industry," *Weld. J.*, **84**(1), pp. 34–40.
- [21] Wang, B., Hu, S. J., Sun, L., and Freiheit, T., 2020, "Intelligent Welding System Technologies: State-of-the-Art Review and Perspectives," *J. Manuf. Syst.*, **56**, pp. 373–391.
- [22] McCauley, R. B., Bennett, M. P., Bodary, W. D., Farrington, G. C., Gasser, R. J., Hurd, W. W., Schueler, A. W., Shearer, T. W., and Silverberg, J. B., 1971, "Resistance Spot Welding," *Metals Handbook, Vol. 6: Welding and Brazing*, T. Lyman, ed., American Society for Metals, Metals Park, OH, pp. 401–424.
- [23] Harkness, J. C., and Guha, A., 1985, "Beryllium-Copper and Beryllium-Nickel Alloys," *Metals Handbook Ninth Edition, Vol. 9: Metallography and Microstructures*, K. Mills, J.R. Davis, J.D. Destefani, D.A. Dieterich, G.M. Crankovic, and H.J. Frissell, eds., American Society for Metals, Metals Park, OH, pp. 392–398.
- [24] Joseph, G., and Kundig, K. J. A., 1999, *Copper: Its Trade, Manufacture, Use and Environmental Status*, ASM International, Materials Park, OH.
- [25] ASTM B 601-02, 2002, *Standard Classification for Temper Designations for Copper and Copper Alloys-Wrought and Cast*, ASTM International, West Conshohocken, PA.
- [26] Zhang, H., 2011, "Evaluation and Quality Control of Resistance-Welded Joints," *Welding Fundamentals and Processes, ASM Handbook, Vol. 6A*, T.J. Lienert, S.S. Babu, T.A. Siewert, and V.L. Acoff, eds., ASM International, Materials Park, OH, pp. 486–504.
- [27] Mansour, T., 1991, "Ultrasonic Testing of Spot Welds in Thin Gage Steel," *Nondestructive Testing Handbook. Vol. 7: Ultrasonic Testing*, P. McIntire, ed., American Society for Nondestructive Testing, Metals Park, OH, pp. 557–568.
- [28] Krautkrämer, J. Krautkrämer, H., 1990, "Welded Joints," *Ultrasonic Testing of Materials*, H. Krautkrämer, J. and Krautkrämer, ed., Springer-Verlag, Berlin, pp. 431–465.
- [29] Martín, Ó., 2004, "Neural Networks Used for the Optimization of Quality Control Processes by Ultrasonics in Resistance Spot Welding Joints (in Spanish)," University of Valladolid.
- [30] Gómez-Ríos, A., Luengo, J., and Herrera, F., 2017, "A Study on the Noise Label Influence in Boosting Algorithms: Adaboost, GBM and XGBoost," *Lect. Notes Comput. Sci.*, **10334 LNCS**, pp. 268–280.
- [31] Bustillo, A., Urbikain, G., Perez, J. M., Pereira, O. M., and Lopez de Lacalle, L. N., 2018, "Smart Optimization of a Friction-Drilling Process Based on Boosting Ensembles," *J. Manuf. Syst.*, **48**, pp. 108–121.

- [32] Freund, Y., and Schapire, R. E., 1996, "Experiments with a New Boosting Algorithm," Proc. 13th Int. Conf. Mach. Learn., pp. 148–156.
- [33] Chen, T., and Guestrin, C., 2016, "XGBoost," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, New York, NY, USA, pp. 785–794.
- [34] Gao, X., Du, L., Xie, Y., Chen, Z., Zhang, Y., You, D., and Gao, P. P., 2019, "Identification of Weld Defects Using Magneto-Optical Imaging," *Int. J. Adv. Manuf. Technol.*, **105**(1–4), pp. 1713–1722.
- [35] Chen, K., Chen, H., Liu, L., and Chen, S., 2019, "Prediction of Weld Bead Geometry of MAG Welding Based on XGBoost Algorithm," *Int. J. Adv. Manuf. Technol.*, **101**(9–12), pp. 2283–2295.
- [36] James, G., Witten, D., Hastie, T., and Tibshirani, R., 2013, *An Introduction to Statistical Learning*, Springer Science & Business Media.
- [37] Breiman, L., 2001, "Random Forests," *Mach. Learn.*, **45**(1), pp. 5–32.
- [38] Gregorutti, B., Michel, B., and Saint-Pierre, P., 2015, "Grouped Variable Importance with Random Forests and Application to Multiple Functional Data Analysis," *Comput. Stat. Data Anal.*, **90**, pp. 15–35.
- [39] Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., and Zeileis, A., 2008, "Conditional Variable Importance for Random Forests.," *BMC Bioinformatics*, **9**(23), p. 307.
- [40] Debeer, D., and Strobl, C., 2020, "Conditional Permutation Importance Revisited," *BMC Bioinformatics*, **21**(1), pp. 1–30.
- [41] Zhou, K., and Yao, P., 2019, "Overview of Recent Advances of Process Analysis and Quality Control in Resistance Spot Welding," *Mech. Syst. Signal Process.*, **124**, pp. 170–198.
- [42] Quinlan, J. R., 1993, *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, San Mateo, CA.
- [43] Rodriguez, J. J., Kuncheva, L. I., and Alonso, C. J., 2006, "Rotation Forest: A New Classifier Ensemble Method," *IEEE Trans. Pattern Anal. Mach. Intell.*, **28**(10), pp. 1619–1630.
- [44] Bagnall, A., Flynn, M., Large, J., Line, J., Bostrom, A., and Cawley, G., 2018, "Is Rotation Forest the Best Classifier for Problems with Continuous Features?"
- [45] Kupervasser, O., 2014, "The Mysterious Optimality of Naive Bayes: Estimation of the Probability in the System of 'Classifiers,'" *Pattern Recognit. Image Anal.*, **24**(1), pp. 1–10.
- [46] Suykens, J. A. K., and Vandewalle, J., 1999, "Least Squares Support Vector Machine Classifiers," *Neural Process. Lett.*, **9**, pp. 293–300.
- [47] Platt, J., 1998, "Fast Training of Support Vector Machines Using Sequential Minimal Optimization," *Advances in Kernel Methods: Support Vector Learning*, B. Schoelkopf, C. Burges, and A. Smola, eds., MIT Press, Cambridge, MA, pp. 185–208.
- [48] Keerthi, S. S., Shevade, S. K., Bhattacharyya, C., and Murthy, K. R. K., 2001, "Improvements to Platt's SMO Algorithm for SVM Classifier Design," *Neural Comput.*, **13**(3), pp. 637–649.

- [49] Üstün, B., Melssen, W. J., and Buydens, L. M. C., 2006, "Facilitating the Application of Support Vector Regression by Using a Universal Pearson VII Function Based Kernel," *Chemom. Intell. Lab. Syst.*, **81**(1), pp. 29–40.
- [50] Hastie, T., Tibshirani, R., and Friedman, J., 2009, *The Elements of Statistical Learning*, Springer, New York, NY.
- [51] Witten, I. H., Frank, E., Hall, M. A., and Pal, C. J., 2016, *Data Mining. Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, Cambridge, MA.
- [52] Finkeldey, F., Volke, J., Zarges, J.-C., Heim, H.-P., and Wiederkehr, P., 2020, "Learning Quality Characteristics for Plastic Injection Molding Processes Using a Combination of Simulated and Measured Data," *J. Manuf. Process.*, **60**, pp. 134–143.
- [53] Wolpert, D., 1992, "Stacked Generalization," *Neural Networks*, **5**, pp. 241–259.
- [54] Ghasemian, A., Hosseinmardi, H., Galstyan, A., Airoidi, E. M., and Clauset, A., 2020, "Stacking Models for Nearly Optimal Link Prediction in Complex Networks," *Proc. Natl. Acad. Sci. U. S. A.*, **117**(38), pp. 23393–23400.
- [55] Prieto, O. J., Alonso-González, C. J., and Rodríguez, J. J., 2015, "Stacking for Multivariate Time Series Classification," *Pattern Anal. Appl.*, **18**(2), pp. 297–312.
- [56] Matlock, K., De Niz, C., Rahman, R., Ghosh, S., and Pal, R., 2018, "Investigation of Model Stacking for Drug Sensitivity Prediction," *BMC Bioinformatics*, **19**(Suppl 3), p. 71.
- [57] Kim, M., Lee, M., An, M., and Lee, H., 2020, "Effective Automatic Defect Classification Process Based on CNN with Stacking Ensemble Model for TFT-LCD Panel," *J. Intell. Manuf.*, **31**(5), pp. 1165–1174.
- [58] Pernía-Espinoza, A., Fernandez-Ceniceros, J., Antonanzas, J., Urraca, R., and Martinez-de-Pison, F. J., 2018, "Stacking Ensemble with Parsimonious Base Models to Improve Generalization Capability in the Characterization of Steel Bolted Components," *Appl. Soft Comput. J.*, **70**, pp. 737–750.
- [59] Ting, K. M., and Witten, I. H., 1999, "Issues in Stacked Generalization," *J. Artif. Intell. Res.*, **10**, pp. 271–289.
- [60] Nadeau, C., and Bengio, Y., 2003, "Inference for the Generalization Error," *Mach. Learn.*, **52**, pp. 239–281.
- [61] Fast, T., Random, U., Ishwaran, A. H., Kogalur, U. B., Kogalur, M. U. B., Suggests, D., Openmp, D. F., Chrome, G., Gpl, L., and Date, R. C., 2021, "Package 'RandomForestSRC'."
- [62] Strobl, C., Boulesteix, A.-L., Zeileis, A., and Hothorn, T., 2007, "Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution," *BMC Bioinformatics*, **8**, p. 25.
- [63] Hothorn, T., 2005, "Survival Ensembles," *Biostatistics*, **7**(3), pp. 355–373.

Figure Captions

- Fig.1. A total of 14 inputs were used to assess the quality level: welding current, welding time, electrode material, treatment applied to electrode material and the ten components of the representative vector of each ultrasonic oscillogram.
- Fig.2. Accuracy of classifiers that obtain at least 90% correct instance classification on the problem depending on the dataset used for training and evaluating the model. In stacking algorithms, the metalearner is specified in parentheses.
- Fig.3. On the left, the ranking of the individual variable importance analysis following Breiman's methodology and on the right the ranking of the individual conditional variable importance in accordance with [39,40,62]. Remarkably, the conditional approach evidences that the welding process variable *welding time* may be the one responsible for the increase in predictive accuracy attained when considering both the welding process parameters and the oscillogram variables.
- Fig.4. Group variable importance in accordance with Gregorutti et al. [38].

Figures

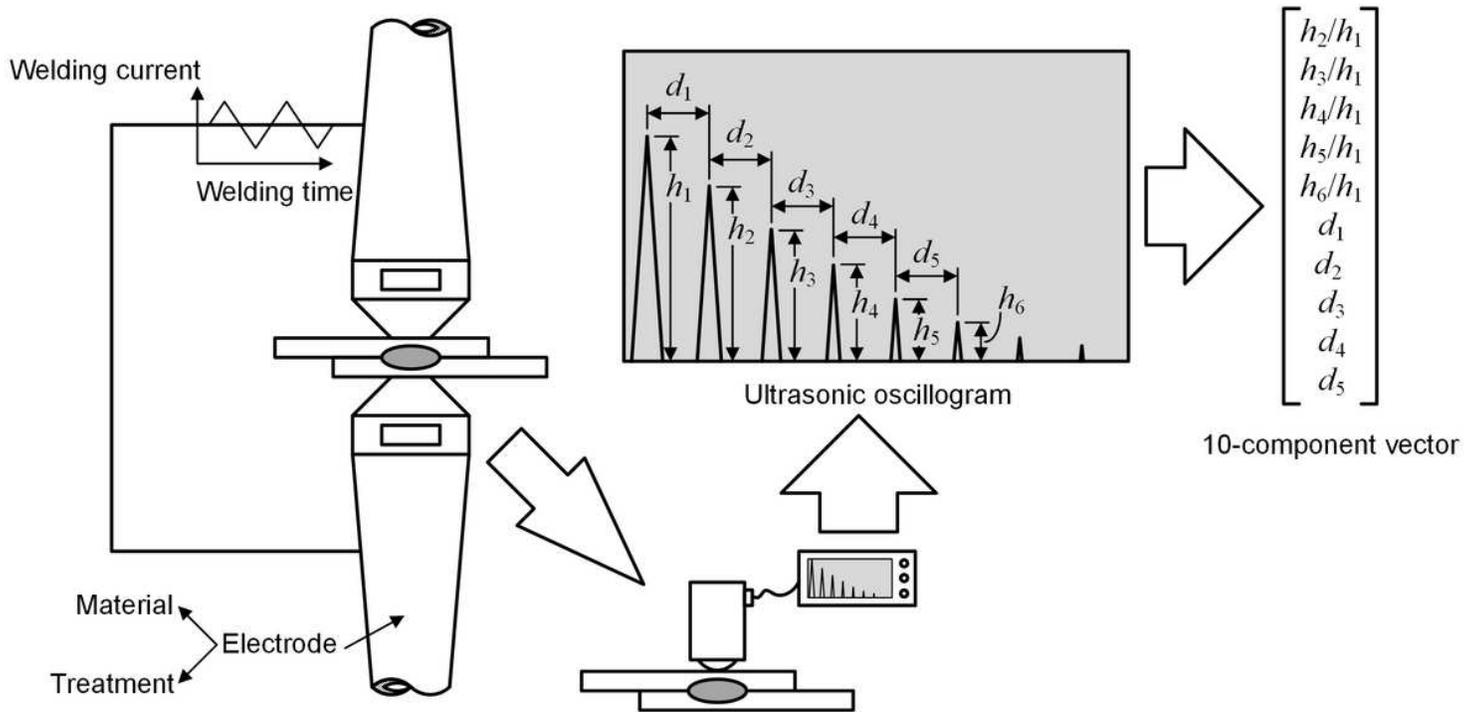


Figure 1

A total of 14 inputs were used to assess the quality level: welding current, welding time, electrode material, treatment applied to electrode material and the ten components of the representative vector of each ultrasonic oscillogram.

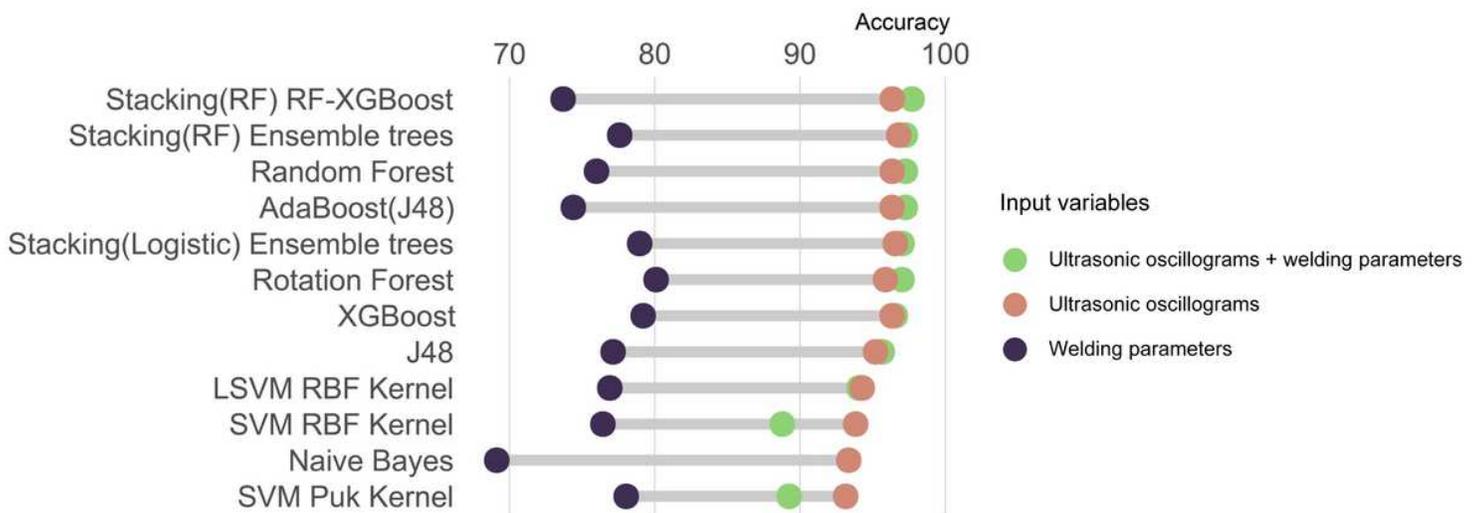


Figure 2

Accuracy of classifiers that obtain at least 90% correct instance classification on the problem depending on the dataset used for training and evaluating the model. In stacking algorithms, the metalearner is specified in parentheses.

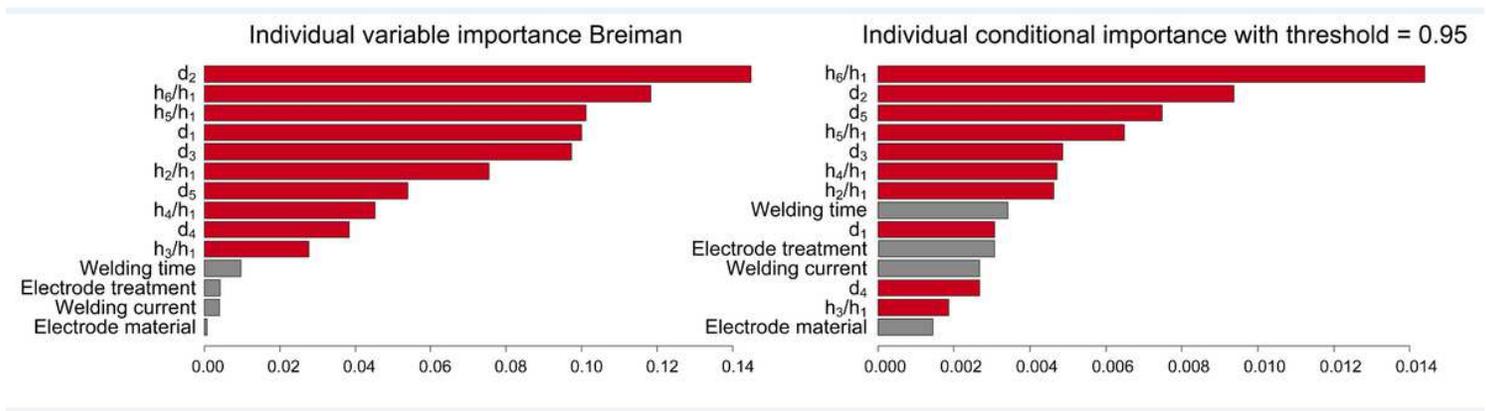


Figure 3

On the left, the ranking of the individual variable importance analysis following Breiman's methodology and on the right the ranking of the individual conditional variable importance in accordance with [39,40,62]. Remarkably, the conditional approach evidences that the welding process variable welding time may be the one responsible for the increase in predictive accuracy attained when considering both the welding process parameters and the oscillogram variables.

Group variable importance

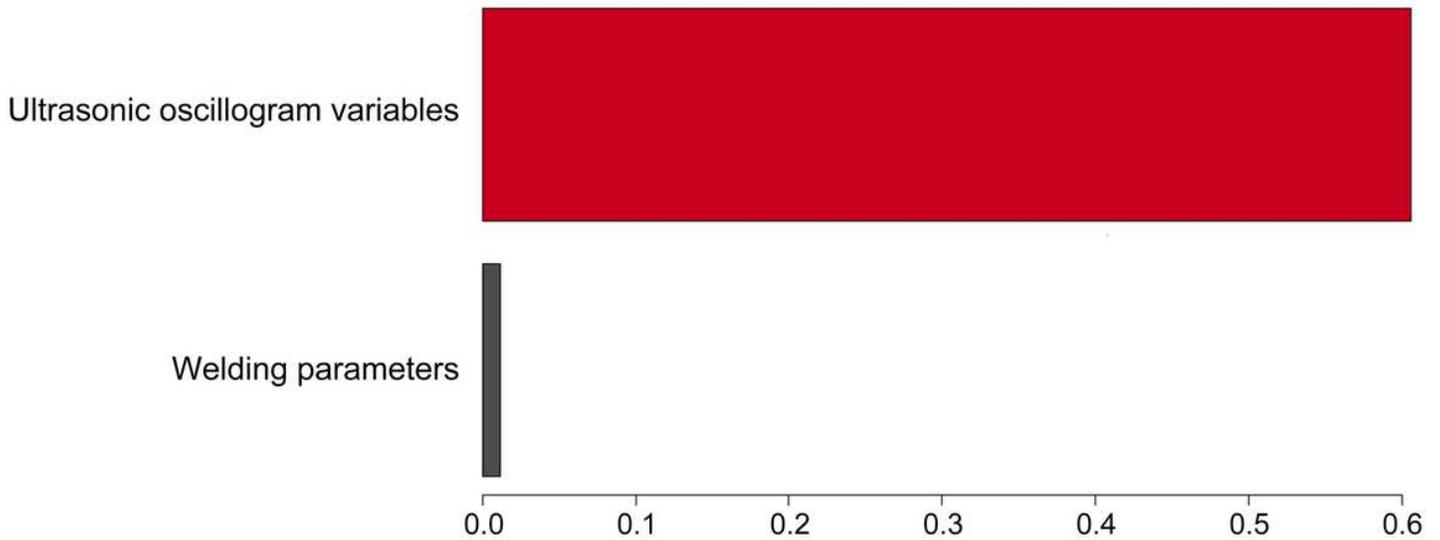


Figure 4

Group variable importance in accordance with Gregorutti et al. [38].