



Glass-box modeling for quality assessment of resistance spot welding joints in industrial applications

José Ignacio Santos¹ · Óscar Martín² · Virginia Ahedo¹ · Pilar de Tiedra³ · José Manuel Galán¹

Received: 21 July 2022 / Accepted: 6 November 2022 / Published online: 18 November 2022
© The Author(s) 2022

Abstract

Resistance spot welding (RSW) is one of the most relevant industrial processes in different sectors. Key issues in RSW are process control and ex-ante and ex-post evaluation of the quality level of RSW joints. Multiple-input–single-output methods are commonly used to create predictive models of the process from the welding parameters. However, until now, the choice of a particular model has typically involved a tradeoff between accuracy and interpretability. In this work, such dichotomy is overcome by using the explainable boosting machine algorithm, which obtains accuracy levels in both classification and prediction of the welded joint tensile shear load bearing capacity statistically as good or even better than the best algorithms in the literature, while maintaining high levels of interpretability. These characteristics allow (i) a simple diagnosis of the overall behavior of the process, and, for each individual prediction, (ii) the attribution to each of the control variables—and/or to their potential interactions—of the result obtained. These distinctive characteristics have important implications for the optimization and control of welding processes, establishing the explainable boosting machine as one of the reference algorithms for their modeling.

Keywords Explainable boosting machine · Pattern recognition · Quality assessment · Resistance spot welding · AISI 304 austenitic stainless steel · Tensile shear load bearing capacity

1 Introduction

Resistance spot welding (RSW) is one of the primary manufacturing methods for joining thin-sheet metal components in the automotive industry [1–4]. Its popularity is mainly due to the significant advantages of the RSW process, such as high welding speed, no need for consumables, high productivity at low cost [5, 6], and the possibility of applying it

on a wide range of base metals—see, for instance, austenitic stainless steels (ASSs), on which RSW is widely used for metro and railway car manufacturing [7–9].

In such a competitive environment as the manufacturing sector, and, more specifically, in the automotive industry, the most outstanding advantage of all the above is precisely its high productivity. Nevertheless, RSW is not without its shortcomings, since, at the same time, it requires a very demanding control of the welding parameters.

From a strictly metallurgical and quality control standpoint, structures with RSW joints are typically devised so that these RSW joints are loaded in shear when the parts are subjected to compression or tension loading [10]. In this regard, the static tensile shear test is the most used—because of its simplicity [11]—to determine the strength of a RSW joint in the laboratory. In particular, the peak load value obtained during the test (the tensile shear load bearing capacity (TSLBC)) is extensively used to estimate the quality of RSW joints [12–14].

As for the most important parameters of the RSW process, these include welding time (WT), welding current (WC), and electrode force (EF) [9, 15], as they significantly affect the

✉ José Manuel Galán
jmgalan@ubu.es

¹ Área de Organización de Empresas, Departamento de Ingeniería de Organización, Escuela Politécnica Superior, Universidad de Burgos, Edificio A1, Avenida Cantabria S/N, 09006 Burgos, Spain

² 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

³ Ciencia de los Materiales e Ingeniería Metalúrgica, Departamento CMeIM/EGI/ICGF/IM/IPF, Universidad de Valladolid, Escuela de Ingenierías Industriales, Paseo del Cauce 59, 47011 Valladolid, Spain

size of the weld nugget which, in turn, crucially influences the TSLBC [16, 17]—i.e., the mechanical properties and quality of RSW joints [18].

It is also important to note that, as pointed out by Zhou and Yao [19], the RSW process is very complex, involving a wide range of interactions between electromagnetic, thermal, mechanical, fluid flow, and metallurgical phenomena. Consequently, recent developments in process analysis and quality control have focused on the obtention of predictive models aimed at both ensuring the integrity of the welded structure and improving welding production efficiency.

Comparative studies on predictive performance on RSW problems generally give black-box models as the top-performing algorithms [19–23]. This type of models provides powerful frameworks for prediction. Still, they often do not offer transparency in their operation, i.e., the rationale behind the choices made by the algorithm is in many cases impenetrable, which, among other problems, makes decision-making and control difficult once an anomalous situation is detected.

Currently, along with the widespread adoption of artificial intelligence-based systems as decision support tools, explainable AI (XAI) has received considerable attention [24, 25]. Traditionally, a tradeoff has been considered to exist between accuracy and interpretability in the use of regression and classification algorithms. Hence, a significant part of XAI methods has been aimed at creating tools that help explain how high-performance black-box models make decisions. Noteworthy, these XAI tools are now challenging the accuracy/interpretability dichotomy in many applications [26]. More precisely, it is a fact that many of the simplest interpretable models that are inherently understandable by humans—such as decision trees or linear regression—do not capture the complexity of many real systems and thus degrade predictive performance compared to black-box models. However, there are other more sophisticated approaches known as glass-box models—specifically designed to be straightforwardly interpretable by humans—that are proving to predict as well—or even better—than the best models to date in many applications, while allowing both global and individual explainability, which may be extremely useful for purposes such as validation and automatic control of manufacturing processes.

In the literature, a wide variety of contributions have addressed quality control and process efficiency in RSW. Martín et al. [27] proposed a neural network-based approach for the prediction of TSLBC in RSW joints using WC, WT, and EF as input factors. Mousavi et al. [28] employed Taguchi

method to design the experiments and Minitab software to analyze the effect of parameters such as WC density, WT, EF, and holding time after welding on tensile-shear strength of the RSW joint. Wang et al. [29] used ultrasonic signal time–frequency features and the PSO-SVM method for RSW joint strength classification. Valaee-Tale et al. [30] presented an analytical model that considered the effects of nugget diameter, EF, base metal yield stress, sheet thickness, and joint fit-up to predict the occurrence of expulsion. Chen et al. [31] used a finite element tool to reveal the evolution of the thermo-electric field and explain the reasons for improved weldability with a slightly concave electrode. Dejans et al. [32] proposed a methodology to reveal linear relationships between nugget diameter and amplitude at a given frequency in the acoustic emission signal. Deng et al. [33] simulated three thermoelectric effects (Peltier, Thomson, and effects) and their impact on asymmetric weld nugget growth in RSW of aluminum. Martín et al. [13] built a regression model implementing polynomial expansion and elastic net regularization of the inputs WC, WT, and EF for TSLBC prediction and quality control classification. Xia et al. [3] employed a multi-sensor monitoring system and a high-speed camera to develop an online RSW expulsion assessment tool from parameters such as instantaneous behaviors in dynamic resistance, EF, and electrode displacement signals.

Since the XAI approach is becoming increasingly popular in many application contexts [34, 35], and given that according to the state-of-the-art analysis in the previous paragraph it has not yet been used for RSW quality control and process improvement, in the present work, we use it to explore the metallurgical problem of TSLBC prediction from the three key RSW parameters WT, WC, and EF. To that end, we chose the glass-box model algorithm known as explainable boosting machine (EBM) [36].

2 Experimental procedure

2.1 Materials and equipment

Table 1 summarizes the chemical composition of the AISI 304 ASS sheets (thickness 0.8 mm) welded by RSW process. Table 2 reports the mechanical properties of the material.

The material was welded using 50-Hz single-phase alternating current (AC) equipment through RWMA group A

Table 1 Chemical composition of the AISI 304 ASS sheets (wt%)

C	Al	Co	Cr	Cu	Mn	Mo	Nb
0.08	0.003	0.17	18.03	0.39	1.153	0.36	0.02
Ni	Si	Ti	W	P	S	V	Fe
8.74	0.426	0.004	0.03	0.019	0.002	0.05	Bal

Table 2 Mechanical properties of the AISI 304 ASS sheets

Tensile strength (MPa)	Yield strength (MPa)	Microhardness (HV, 100 g)	Total elongation (%)
675	290	162	70

class 2 water-cooled truncated cone electrodes with a face diameter of 4.5 mm.

2.2 Obtaining the tensile shear test specimens by RSW

The welding parameters controlled in the RSW process were—because of their proven relevance—[13, 27, 37] (i) WT—which varied from 0.24 to 0.04 s, with a 0.02 s step decrease; (ii) WC—which varied from 6.5 to 1.5 kA RMS (approximately) with a 0.5 kA RMS step decrease; and (iii) EF—which was tested at two different values: 1000 and 1500 N.

The experiments were set up to sweep the welding parameter ranges described above. In particular, 242 ($2 \times 11 \times 11$) different possible welding configurations were considered, being a tensile shear test conducted on each spot-welded specimen. Recall that spot welding was conducted in accordance with the procedure described in an earlier work [27] and following the ISO 14273 standard [38].

The weld nugget of a RSW joint has a characteristic shape with an as-cast dendritic microstructure. Figure 1 shows different macro- and micrographs of RSW specimens obtained with different process parameters. More precisely, Fig. 1(A–D) corresponds to a high-heat input RSW joint obtained with a WT of 0.22 s, a WC of 5.51 kA, and an EF of 1500 N, and whose TSLBC—from the tensile shear test—is 6.44 kN. On their part, Fig. 1(E–H) corresponds to a low-heat input RSW joint obtained with a WT of 0.08 s, a WC of 3.41 kA, and an EF of 1500 N, and whose TSLBC—from the tensile shear test—is 3.78 kN.

Notably, Fig. 1 illustrates two relevant phenomena:

- (i). The decisive influence of the heat input on the size of the weld nugget and, consequently, on the TSLBC. This can be clearly seen by comparing the high-heat input RSW joint from Fig. 1(B)—which exhibits a large weld nugget and a TSLBC of 6.44 kN—to the low-heat input RSW joint from Fig. 1(F)—whose weld nugget is significantly smaller and whose TSLBC is 3.78 kN.
- (ii). For the considered values of the welding parameters, no significant differences in grain size are observed, neither in the heat-affected zone (HAZ) nor in the weld nugget. In fact, comparison of grain size between Fig. 1(A)—HAZ of a high-heat input

RSW joint obtained with WT = 0.22 s and WC = 5.51 kA—and Fig. 1(E)—HAZ of a low-heat input RSW joint obtained with WT = 0.08 s and WC = 3.41 kA—shows no significant differences. As for the weld nugget area, the differences in grain size between Fig. 1(C, D)—weld nugget of the high-heat input RSW joint—and Fig. 1(G, H)—weld nugget of the low-heat input RSW joint—are not significant either. This suggests that, to observe the increase in grain size with increasing heat input in both the HAZ and the weld nugget, as demonstrated by [10], higher WCs than those used in our experiments may be required—recall that WC is the key parameter in the heat generated by Joule effect.

2.3 Quality assessment from TSLBC values

Tensile shear tests were conducted on the 242 RSW specimens at a crosshead speed of 2 mm/min, which allows considering the test as static [40, 41]. As a result, 242 TSLBC values were obtained.

JIS Z 3140 standard [42] sets the minimum value of the weld nugget diameter at 4.5 mm (when the sheet thickness is equal to 0.8 mm) to consider a RSW joint acceptable. By means of ultrasonic testing, we checked for this requirement on all samples. The transducer diameter selected for the ultrasonic test was equal to the aforementioned acceptance threshold [43], i.e., 4.5 mm. The transducer (with a frequency of 20 MHz) operated with a captive water column delay and a replaceable rubber membrane that facilitated coupling with the RSW joint surface. Importantly, the oscillogram of an acceptable RSW joint is characterized by a short echo sequence span, as the coarse as-cast dendritic microstructure of the weld nugget (Fig. 1) increases attenuation. Note in Fig. 2 that the distance between consecutive echoes in an acceptable RSW joint is the combined thickness of the two welded sheets, since the ultrasonic beam reflections take place at the lower sheet's bottom surface. On the other hand, an unacceptable RSW joint presents both principal echoes of reflections occurring at the bottom surface of the lower sheet—which are produced by the portion of the ultrasonic beam that goes through the weld nugget—and one-layer echoes between the principal ones that are due to reflections at the interface between sheets—which are produced by the part of the ultrasonic beam that does not go through the weld nugget (Fig. 2) [13, 44].

In the shear tests described above, the minimum TSLBC obtained for a valid/acceptable RSW joint was 5.93 kN. Therefore, the minimum admissible TSLBC was set at 5.93 kN (Fig. 2), which is even more conservative than the criterion also established by the JIS Z 3140 standard [42] based on the TSLBC itself that, for a 0.8-mm sheet thickness (considering that the tensile strength of the base metal was above

Fig. 1 **A** Micrograph showing the HAZ and the weld nugget (WN) in a high-heat input RSW joint. **B** Macrograph showing the weld nugget (WN) size of a high-heat input RSW joint. **C, D** Micrographs showing the as-cast dendritic microstructure of the weld nugget of a high-heat input RSW joint—**D** at higher magnification than **C**. **E** Micrograph showing the HAZ and the weld nugget (WN) in a low-heat input RSW joint. **F** Macrograph showing the weld nugget (WN) size of a low-heat input RSW joint. **G, H** Micrographs showing the as-cast dendritic microstructure of the weld nugget of a low-heat input RSW joint—**H** at higher magnification than **G**. Electrolytic etching with oxalic acid according to ASTM A 262–91 Practice A [39]

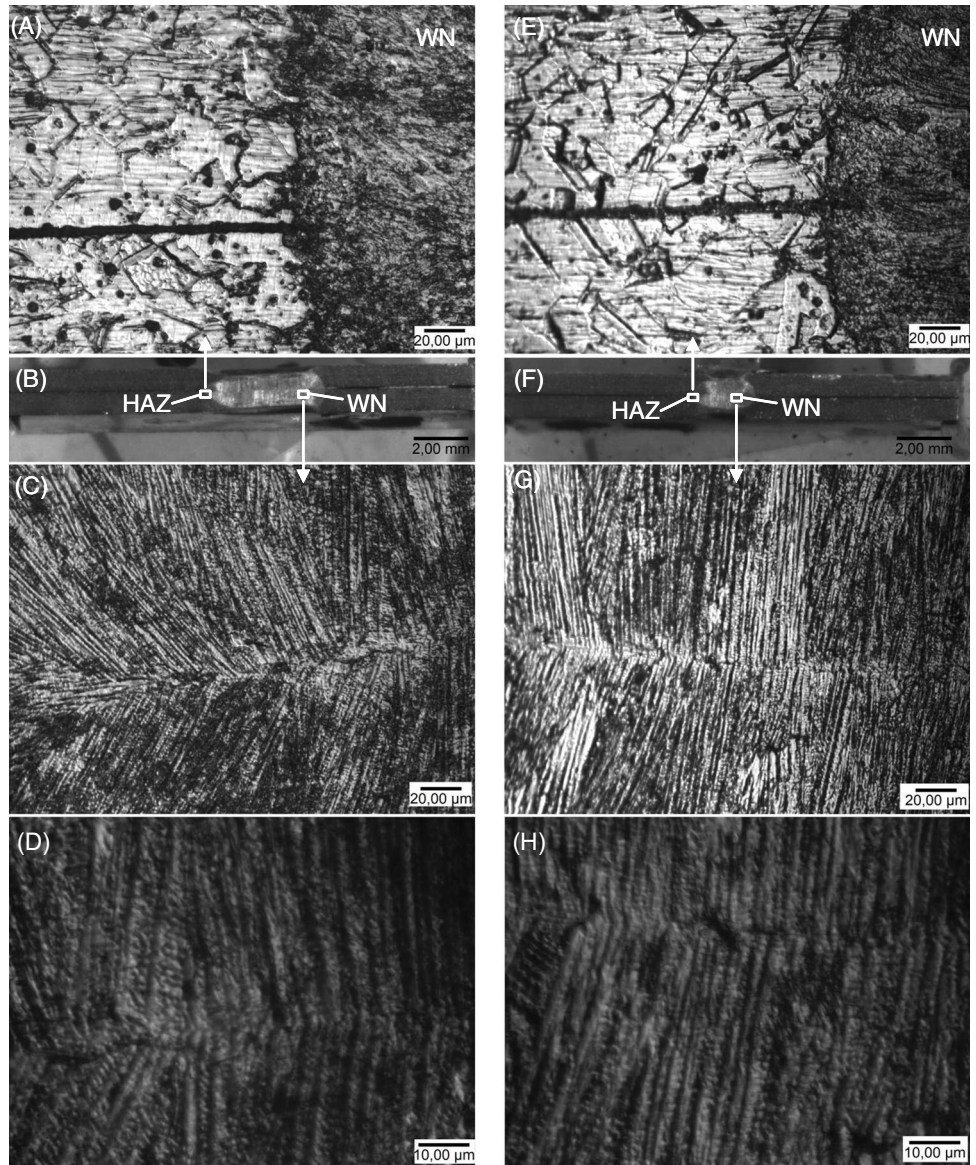
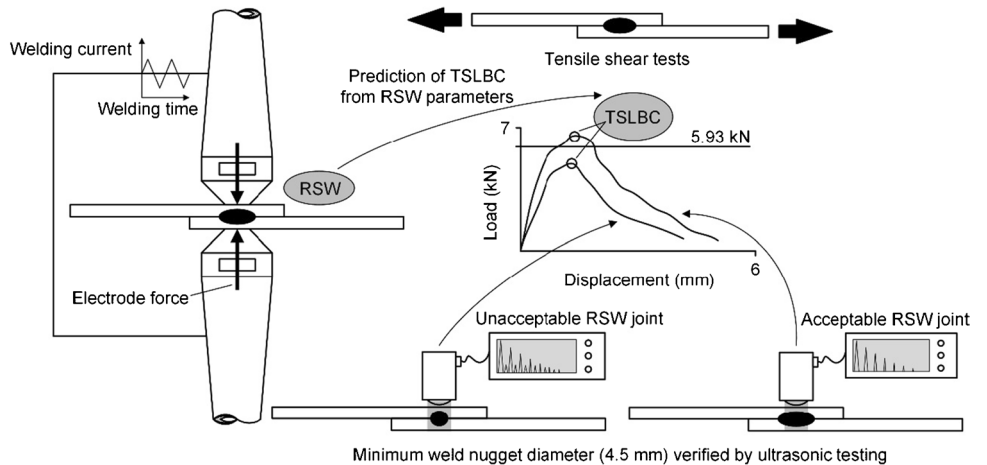


Fig. 2 Diagram of the procedure for assessing the quality of RSW joints from TSLBC values



590 MPa), sets the minimum TSLBC at $3.53 \text{ kN} \times 1.6 = 5.65 \text{ kN}$.

3 Theoretical background

3.1 Explainable boosting machine

The EBM [36] is a glass-box model specifically designed to obtain high predictive performance, comparable or superior to state-of-the-art machine learning methods, while being intelligible and explainable.

The EBM is inspired by the generalized additive models (GAMs) [45]. GAMs have the following structure:

$$g(E[y]) = \beta_0 + \sum_j f_j(x_j) \quad (1)$$

where the model relates a response variable y to be predicted—assumed to follow some distribution of the exponential family—with a set of predictor variables x_j following the structure of Eq. (1). Note that g is a link function capable of flexibly adapting to the type of outcome required—typically different in regression and classification problems. In contrast to generalized linear models (GLMs), GAMs model the response variable as a sum of arbitrary univariate functions that are not necessarily linear. In other words, the response is assumed to be a sum of effects by features, where the relationship of each variable with the output may be nonlinear. Typically, in GAMs, each function f_j is modeled through spline functions.

The EBM is an efficient implementation of the GAM plus the interactions (GA²M) algorithm [46], which allows multi-core and multi-machine parallelization [36]. The GA²M algorithm includes basically two features to improve performance over GAM: (i) f_j functions are not modeled through splines but through tree ensemble-based techniques such as bagging and gradient boosting, and (ii) GAM models are generalized to allow the detection of pairwise relevant interactions between variables and their inclusion in the predictive model (see Eq. (2)).

$$g(E[y]) = \beta_0 + \sum_j f_j(x_j) + \sum_{i \neq j} f_{ij}(x_i, x_j) \quad (2)$$

Both the inclusion of interaction terms in EBM models and the use of tree ensemble-based univariate functions instead of splines significantly increase their performance [36, 47–52]. Moreover, EBMs have the additional advantage that they are still interpretable. Through the visualization of the f_j functions and the interaction terms, it is possible to reason about the contribution of each feature to the final prediction, both in isolation and in relation to its potential interactions with other variables—in case they exist and are relevant.

The fitting process of the f_j functions is performed by restricting the training set to one feature at a time through a round-robin procedure with a very low learning rate, and by running boosting iterations over the dataset thousands of times. This approach renders irrelevant the order in which the features are chosen in the training process and reduces the effects of possible collinearity between variables. In addition, the use of gradient boosting with ensembles of shallow regression trees has been shown experimentally to give higher accuracy than the spline functions used in traditional GAM fitting [53].

The process of detecting interactions between variables in high-dimensional datasets can be expensive in both computational and interpretation terms. Therefore, the GA²M algorithm only includes those interactions that pass a given statistical test. Specifically, it uses an efficient algorithm, called FAST [46], to establish a ranking of the strength of the interactions; from the ranking, a greedy forward selection strategy is used until convergence, being selected in the model only those interactions that are relevant for prediction.

As we illustrate in our case study below, the EBM has implementations for both regression and classification, and provides two frameworks for interpretation: the global approach and the case-level one. As regards the global approach, the model allows for an additive decomposition of the response into easily understandable terms once the model is fitted. For example, one can determine (i) how each variable contributes to the final prediction through one-dimensional feature functions, (ii) the impact of the relevant interaction pairs—if they exist—using two-dimensional functions, and (iii) the global estimates of the importance of each predictor in the model. From the case-level perspective, the above decomposition is performed at the level of each individual instance, thus allowing to identify the variables that were determinant for a specific prediction through their contribution scores. This aspect is a fundamental advantage in controlling manufacturing processes, since it sheds light on the cause of the deviations that led to a given result.

3.2 Computational experiments and description of the other regression and classification algorithms selected

Our computational experiments were structured around two main approaches: regression and classification. In both cases, the performance of the EBM model was compared with that of the most relevant set of algorithms according to the literature for the problem of quality control from welding parameters—which were used as a performance baseline [19, 21–23, 54]. Remarkably, in the classification approach, RSW joints were classified as valid if their TSLBC was equal to or greater than 5.93 kN, and as invalid otherwise. Please note that before training the algorithms, the dataset

was standardized (to obtain zero mean and unit variance in each variable), because some algorithms—e.g., support vector machines—are sensitive to the distribution of values.

Most regression and classification algorithms require choosing the values of a set of parameters known as tuning parameters or hyperparameters. In the present work, once the tuning parameters were chosen and the different models fitted, their comparison was conducted based on their test errors. To reduce bias in the estimation of the latter, nested cross-validation (NCV) was implemented [55, 56]. NCV involves applying cross-validation (CV) through two sequential loops: an inner loop for selecting the hyperparameters and an outer loop for computing the test error. In our experiments, both loops—inner and outer—implemented tenfold CV. In the classification problem, due to an imbalance in the distribution of the two classes—the dataset contains significantly more acceptable than unacceptable RSW joints—stratified nested CV—a refinement specifically conceived to address the imbalance problem—was used. As for the evaluation metrics, regression models were compared using the mean squared error (MSE), and classification models by means of the classification error (complementary of the accuracy).

Specifically, the algorithms used were four ensemble algorithms (combination of weak learners to obtain a strong learner) based on decision trees (XGBoost, AdaBoost, gradient boosting machines (GBMs), and random forests (RFs)), support vector machines (SVMs), deep learning models (DLs), GAMs using smoothing splines), and linear regression with elastic net regularization. For algorithms with both regression and classification versions, the appropriate one was chosen for each case. The following is a summary description of the algorithms considered.

Boosting is one of the most widely used ensemble techniques for both regression and classification due to its high performance in many domains. Boosting consists of obtaining a strong classifier from the sequential combination of weak base learners. There are different alternatives and implementation nuances among the algorithms that use boosting as a primary construction mechanism [57]. In this work, three different boosting algorithms were used: AdaBoost [58], GBM [59], and XGBoost [60]. More precisely, AdaBoost was built using as weak learner three-level depth decision trees for regression and decision stumps for classification. Basically, in each iteration, the process consists of reweighting the training set so that those data that were misclassified in previous runs receive a higher weight in subsequent iterations. The algorithm also estimates the learning parameter that weights the contribution of each new learner in the process. In the GBM algorithm, boosting is implemented so that, at each iteration, the latest weak learner (deep regression trees for both regression and classification) is trained on the residuals obtained in previous iterations.

The training process to minimize the loss function expectation is performed iteratively using a stochastic gradient descent scheme. Eventually, the XGBoost algorithm—like the GBM—consists of an additive expansion of regression trees using boosting. However, compared to the GBM, it incorporates different mechanisms that enable parallelization and more efficient memory use, thus reducing computation times. Remarkably, although this is an algorithm originally designed for significantly larger datasets than the one in this work, previous experiments have also shown good results on small and medium-sized datasets of engineering problems in general [61–63] and welding processes in particular [21, 64, 65].

In contrast to boosting, random forests [66] use decision trees as well, but in this case with a different ensemble technique: bootstrap aggregation (bagging) [67]. Bagging consists of combining the results obtained by different weak learners to later aggregate them by means of—generally—averaging in regression and a voting system in classification. In addition to bagging, RFs also use the random subspace method to decorrelate trees to each other. RF is considered one of the best out-of-the-box algorithms due to its good predictive performance in many application contexts and its robustness to overfitting and possible internal correlations between explanatory variables.

In addition to ensemble approaches, we used other state-of-the-art high-performing algorithms for regression and classification, i.e., SVMs, DL, GAMs, and linear models, with elastic net regularization.

SVMs [68] are very popular and robust regression and classification algorithms with excellent generalization performance. They are based on constructing a hyperplane—or a set of hyperplanes—by maximizing the margin of separation between classes. When classes are not linearly separable, the original dataset is mapped into a higher-dimensional space using kernel functions.

DL is a machine-learning approach based on multilayer artificial neural networks widely used for its universal approximation capabilities [69]. DL models consist of several interconnected layers of units, called neurons. A neuron is a non-linear function, i.e., an activation function, which takes as input the weighted sum of the outputs of neurons in the previous layer (plus a constant offset) and passes its output to the neurons of the following layer. Typically, DL models implement many hidden layers. In this work, we used the Fastai python library, which offers a friendly API for using the popular deep learning framework PyTorch [70]. Remarkably, Fastai specifically provides an architecture for working with tabular data (as our experimental dataset), facilitates the work with categorical and continuous variables, and the training of multilayer models (i.e., weights and offsets of neurons) using the stochastic gradient descent algorithm [71].

GAMs are generalized linear models that model the variable to be predicted as the sum of dependent functions of the regressors (along with a link function chosen for the type of problem) [45]. In GAMs, unlike EBM, these functions are often regression splines.

Finally, we have included linear models with elastic net regularization [72] given their good predictive performance in the RSW quality prediction problem [13]. Regularization is the most commonly used technique to overcome overfitting problems in linear regression. As for elastic nets, they constitute a specific regularization scheme that allows weighting between Lasso and ridge regularization. Therefore, elastic nets address both the problem of obtaining parsimonious models (variable selection) and the possible correlations between predictors.

4 Results and discussion

4.1 Comparative analysis for regression

Table 3 presents the results obtained for the regression problem of predicting the TSLBC from the key welding parameters WT, WC, and EF. The performance metric chosen was the MSE, which was calculated by means of tenfold nested

CV. As previously stated, the performance of the regression EBM is compared to that of the best-performing regression algorithms in the literature.

The interpretation of Table 3 is as follows: the regression explainable boosting machine (MSE=0.031 and std=0.013) is the best performing algorithm for TSLBC prediction on the RSW dataset under consideration. Notwithstanding, it is closely followed by the regression gradient boosting machine (MSE=0.033 and std=0.015), the random forest for regression (MSE=0.034 and std=0.017), and the Regression XGBoost (MSE=0.036 and std=0.015).

4.2 Comparative analysis for classification

Table 4 summarizes the results obtained for the classification problem of determining whether a RSW joint is valid or invalid. As previously explained, we defined these two classes in accordance with the TSLBC: a RSW is valid if its TSLBC ≥ 5.93 kN, and invalid otherwise. The performance metric selected was the classification error, which was obtained by tenfold stratified nested CV. In this case, the performance of the classification EBM is compared to that of the best-performing benchmark classification algorithms.

The interpretation of Table 4 is as follows: in our RSW dataset, the best classification algorithm is the

Table 3 Summary table of the results obtained for the regression approach. Regression algorithms sorted from lowest to largest mean MSE

	MSE	
	Mean value	Standard deviation
Explainable boosting machine (ExplainableBoostingR)	0.031	0.013
Gradient boosting machine (GradientBoostingR)	0.033	0.015
Random forest (RFR)	0.034	0.017
XGBoost (XGBoostR)	0.036	0.015
GAM	0.039	0.011
Deep learning	0.042	0.017
AdaBoost (AdaBoostR)	0.042	0.024
Support vector machine (SVR)	0.051	0.029
Linear regression with EN regularization (linear-EN)	0.118	0.049

Table 4 Summary table of the results obtained for the classification approach. Classification algorithms sorted from lowest to largest classification error

	Classification error	
	Mean value	Standard deviation
Deep learning	0.0658	0.0393
AdaBoost (AdaBoostC)	0.0658	0.0584
Explainable boosting machine (ExplainableBoostingC)	0.0697	0.0539
Random forest (RFC)	0.0697	0.0502
Log-GAM	0.0698	0.0507
Support vector machine (SVC)	0.0742	0.0252
Gradient boosting machine (GradientBoostingC)	0.0780	0.0589
XGBoost (XGBoostC)	0.0822	0.0608

deep learning one (mean classification error = 0.0658, std = 0.0393), followed by the classification AdaBoost (mean classification error = 0.0658, std = 0.0584) and the classification explainable boosting machine (mean classification error = 0.0697, std = 0.0539).

4.3 Frequentist statistical tests

The traditional approach to compare the performance of different algorithms consists of applying classic frequentist tests to their final results; this is known as the null hypothesis statistical test (NHST) paradigm. Nevertheless, in recent years, increasing criticisms have emerged against the frequentist approximation, being Bayesian tests proposed to overcome some of their limitations [73].

In the frequentist realm, two main types of tests are the most common: parametric and non-parametric tests. In both cases, a non-effect null hypothesis is assumed, and a NHST is performed. Hence, just a single probability value is computed and compared against the confidence coefficient of choice.

In the present contribution, the corrected paired Student *t*-test was selected since, as stated in [74], it accounts for both the variability due to the choice of the training sets and that of the test sets, thus being less biased than ordinary tests of significance, and having the correct size and power.

4.3.1 Frequentist statistical analysis for regression

The pairwise results of the paired Student *t*-test calculated on the tenfold cross-validation results of the different regression algorithms selected are presented in Fig. 3. A significant threshold value of 0.05 was selected. *P*-values lower than 0.05 are represented in green, and those above it are represented in red. Statistically significant differences exist between the EBM and XGBoost, GAM, DL, SVR, and linear regression, which may be interpreted as the EBM generalizing better than these state-of-the-art algorithms.

4.3.2 Frequentist statistical analysis for classification

The pairwise results of the paired Student *t*-test calculated on the tenfold stratified nested CV results of the different classification algorithms selected are summarized in Fig. 4. The color scale used in Fig. 4 is the same as in Fig. 3. As it can be seen, for classification, there are no statistically significant differences between the different algorithms considered, at least for the level of significance chosen (0.05).

4.4 Bayesian statistical analysis

A Bayesian statistical analysis was also conducted to complete the comparative analyses in the previous section (null hypothesis statistical tests defined to determine the significance levels of Figs. 3 and 4). In the Bayesian approach,

Fig. 3 *P*-values obtained for the regression problem using the corrected paired Student *t*-test (frequentist approach)

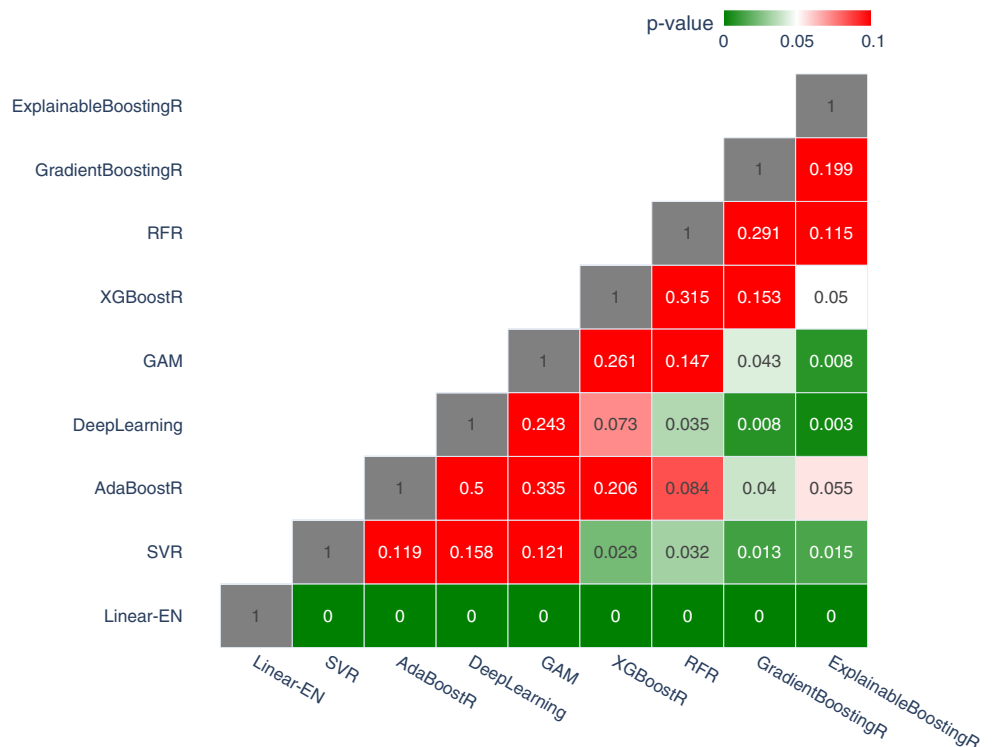
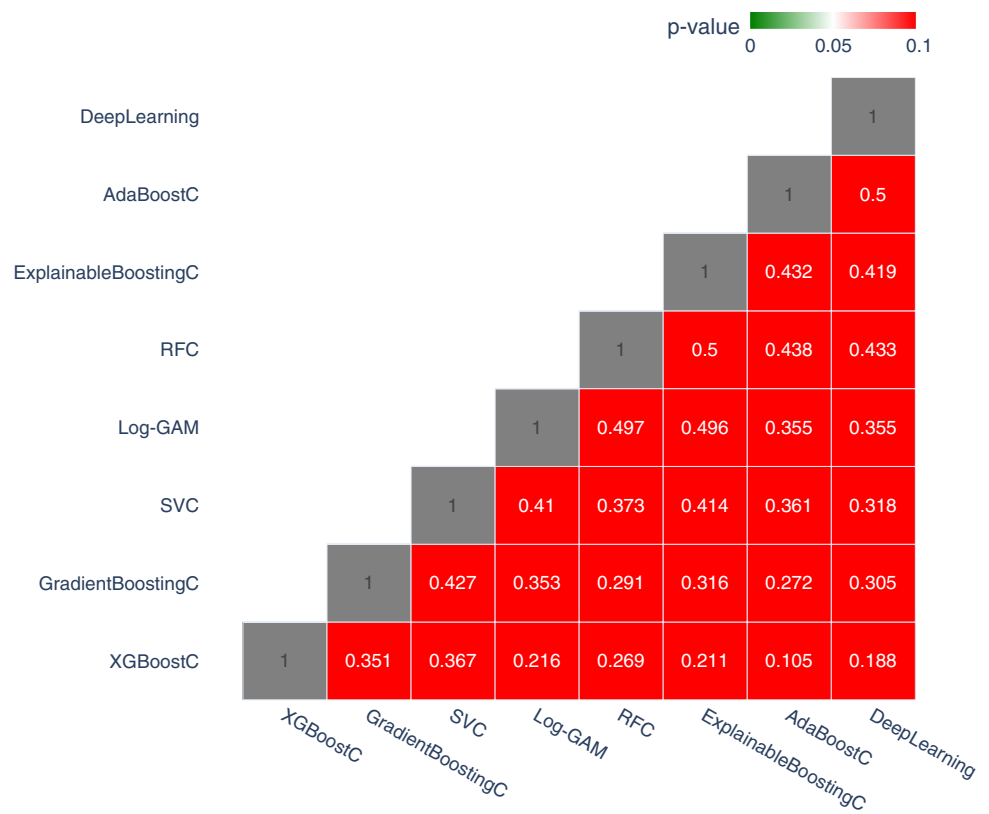


Fig. 4 *P*-values obtained for the classification problem using the corrected paired Student *t*-test (frequentist approach)

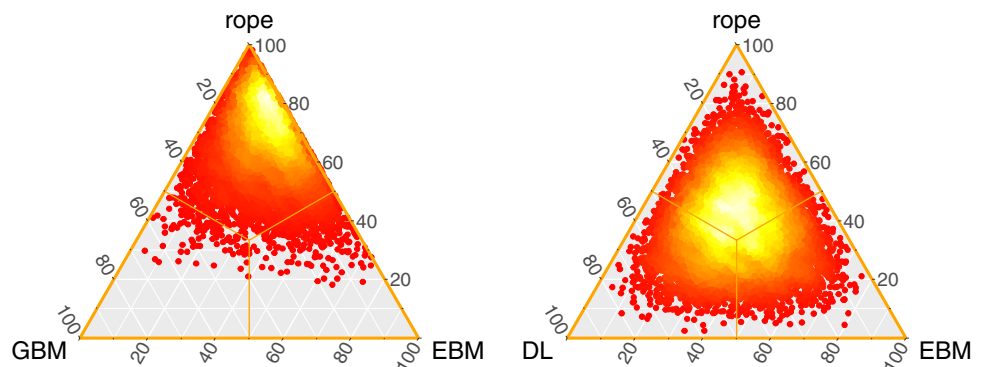


a probability distribution is established on the parameter of interest. This type of test is easier to interpret since it estimates the probabilities of the hypotheses, allowing to distinguish between magnitude and uncertainty, among other advantages [73, 75]. Specifically, we used the Bayesian correlated *t*-test for cross-validation on a single dataset to compare, on the regression approach, the differences between the two empirically found best-performing algorithms—regression explainable boosting machine and regression gradient boosting machine. On its part, for classification, we used the same test and compared the best-performing classifier—the deep learning classifier—to our explainable algorithm of interest: the classification

explainable boosting machine—which happens to be the third best classifier on our dataset.

In the Bayesian framework, a region of practical equivalence (rope) can be defined: some limits of comparison for which we consider that the magnitude of the effect found is not relevant. In regression, the area of relatively no effect was defined with a limit difference of 0.005 in the MSE, which is approximately sensitive to relevant variations in the prediction of the first decimal place of the TSLBC. In the classification problem, the area of practically no effect was established for differences in accuracy of less than 1%—the most common threshold value in the literature.

Fig. 5 Bayesian correlated *t*-test results for the two (empirically found) best algorithms for regression (left) and for the best classifier and the EBM in the classification approach (right). In both cases, three probabilities were estimated and plotted in a ternary graph: the probability that one of the algorithms is better than the other (left and right in each graph), and the Region of Practical Equivalence (rope) (above)



The results obtained are shown in Fig. 5. In the regression case (left in Fig. 5), the results show that the probability that the EBM is the best algorithm is 21%, the probability that it is the GBM is 12%, and that the probability that the differences are in the range of the rope is 67%. In the classification case (right in Fig. 5), with the available evidence, the probability that Deep Learning is the best algorithm is 30%, that it is the EBM is 30%, and that there is no difference above the rope limits is 40%. The results from the Bayesian analysis—for both regression and classification—confirm again that the EBM has a very high probability of being either the best algorithm or to have a difference with the best algorithm that lies in the equivalence region—i.e., that it can be considered equivalent to the best algorithm. Therefore, from the results obtained with both statistical approaches, the conclusion to be drawn is that the use of the EBM does not statistically degrade the predictive accuracy (since it is either the best algorithm or equivalent to the best algorithm), while having the advantage of being interpretable, as we shall see in detail in the next section.

4.5 Global results for EBM and interpretation

The results of the analysis of the overall importance of the welding parameters (Fig. 6) show that the two variables that

most influence the TSLBC are WC and WT, with EF and each of the pairwise interactions between the three welding parameters considered having a marginal influence on the TSLBC.

Also noteworthy is the outstanding importance of WC with respect to the rest of the variables analyzed, as it is 3.7 times more important than the second most important variable, which is WT. This result is consistent with the fact that WC is the only quadratic term in the Joule heating equation, which is what gives rise to the formation and growth of the weld nugget [13]. Therefore, since weld nugget size is the most critical determinant impacting TSLBC [16, 17], and WC is the most important parameter affecting weld nugget size [76, 77], WC is the most relevant parameter affecting the TSLBC.

Figures 7 and 8 clearly and intuitively represent the expected relationship between the welding process control variables and the response. They show that the average TSLBC value is increased or decreased as a function of the values of WC and WT. Figure 9 also shows that there is an interaction between WC and WT. This interaction slightly modulates the effect of the primary variables and increases the accuracy of the model. There is also a step effect that depends on the EF used in the experimental phase; nevertheless, its figure has not been

Fig. 6 Overall importance (mean absolute score) of the different predictor variables considered in the EBM regression model built to predict the TSLBC. Recall that the EBM considered pairwise interactions between variables

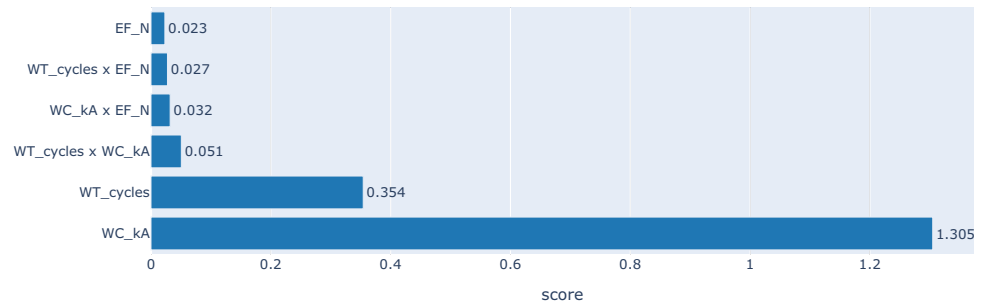


Fig. 7 Increase or decrease of the TSLBC with respect to the average TSLBC (score) of the dataset as a consequence of the individual increase in WC—WC on the abscissa axis

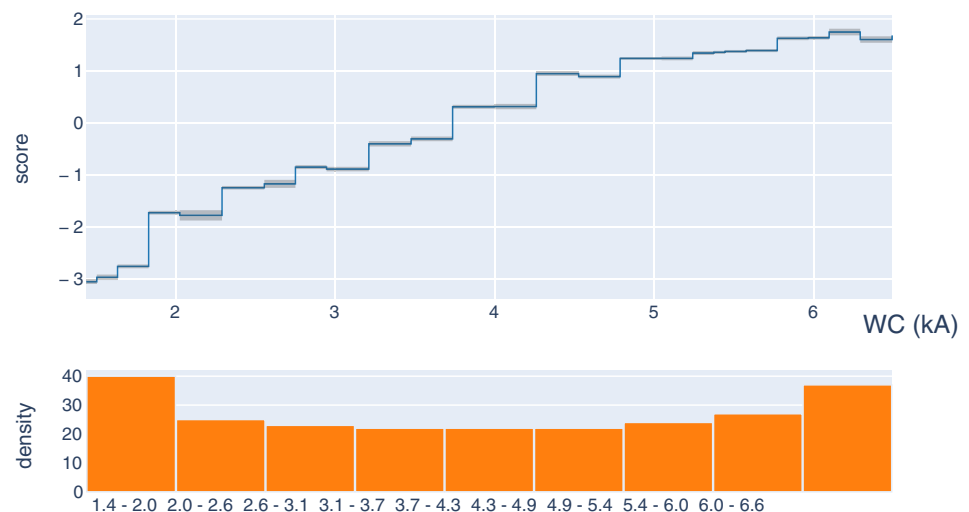


Fig. 8 Increase or decrease of the TSLBC with respect to the average TSLBC (score) of the dataset as a consequence of the individual increase of WT—WT on the abscissa axis

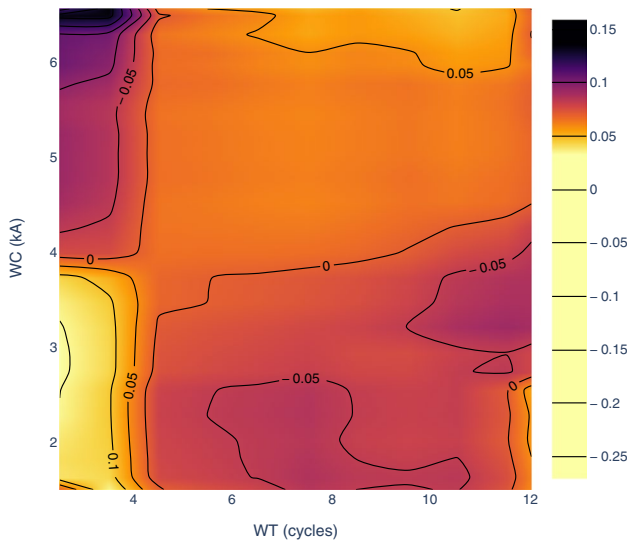
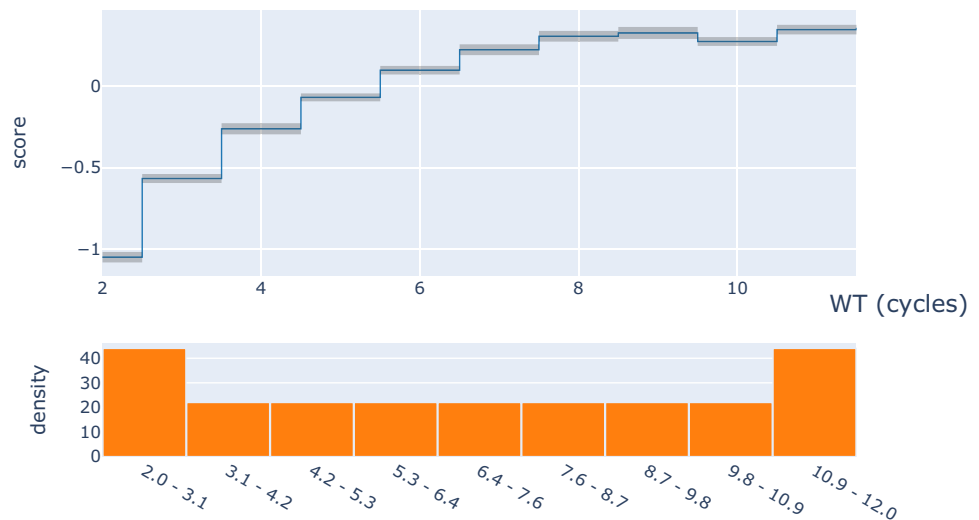


Fig. 9 Contour plot of the pairwise interaction between WT and WC and its effect on the TSLBC

included for simplicity. The results show that, in general, as WC and WT increase—and thus heat input—TSLBC also increases. However, when the heat input is excessive, the expulsion phenomenon occurs and, as a result, the TSLBC decreases [78, 79]. Note that, for high WTs, and irrespective of whether WC is high or low (i.e., for the average of all values of WC), increases in WT do not lead to significant increases in TSLBC (Fig. 8). However, if the focus is on the WC, and, more specifically, on the region of high WCs in Fig. 7, irrespective of whether WT is high or low (i.e., for the average of all values of WT), increases in WC lead to decreases in TSLBC. This behavior is explained by the expulsion phenomenon—which

is strongly influenced by the WC because of its impact on the heat generated by Joule effect as previously noted [18, 80–82]).

4.6 Analysis of individual results

As previously stated, the EBM algorithm allows for a case-level analysis in which, for each individual RSW joint, it is possible to determine the contribution of the value of each welding variable to the TSLBC obtained. This can be useful, for example, in fault diagnosis, to attribute a predicted TSLBC value of less than 5.93 kN to one or more welding variables.

To illustrate the utility of the EBM as a fault diagnosis tool, two RSW joints with predicted TSLBC values of less than 5.93 kN are studied. Recall that the predicted TSLBC value is the sum of the values/contributions of each predictor variable (the intercept being the average TSLBC value in the training set).

Figure 10 shows, for the first RSW joint studied, that the welding variable that “pulls down” the final predicted TSLBC value—and therefore the main responsible for the RSW joint being invalid—is the WT (which contributes by subtracting a value of 0.567 kN).

Figure 11 shows, for the second RSW joint studied, that the welding variable that “pulls down” the final predicted TSLBC value—and therefore the main responsible for the invalid status of the RSW joint—is the WC (which contributes by subtracting a value of 0.847 kN).

By way of summary, in our application example, the EBM has obtained—both for regression and classification—a predictive performance as high as that of the best black-box algorithms to date, while allowing interpretation of the results obtained. In this regard, we have also illustrated its interest and potential both for the global analysis of the RSW process (process-control tool) and for case-level insights (fault diagnosis tool). Therefore,

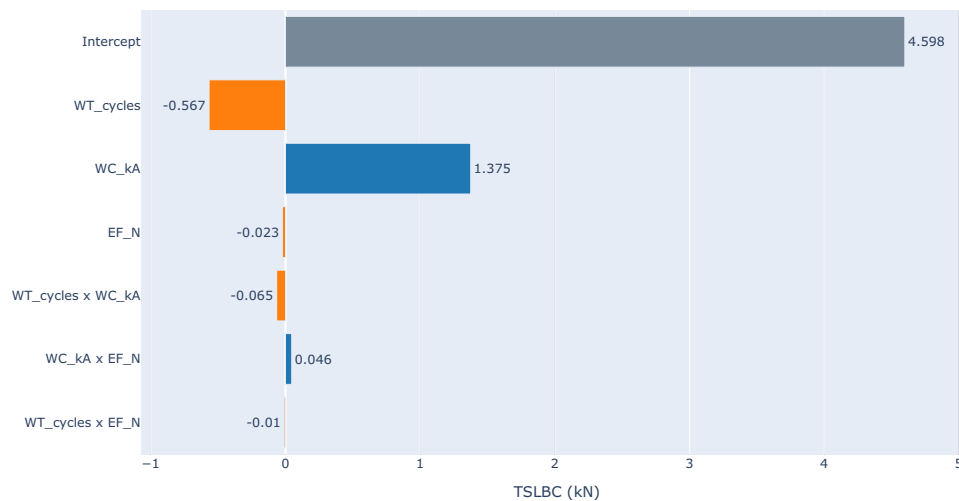
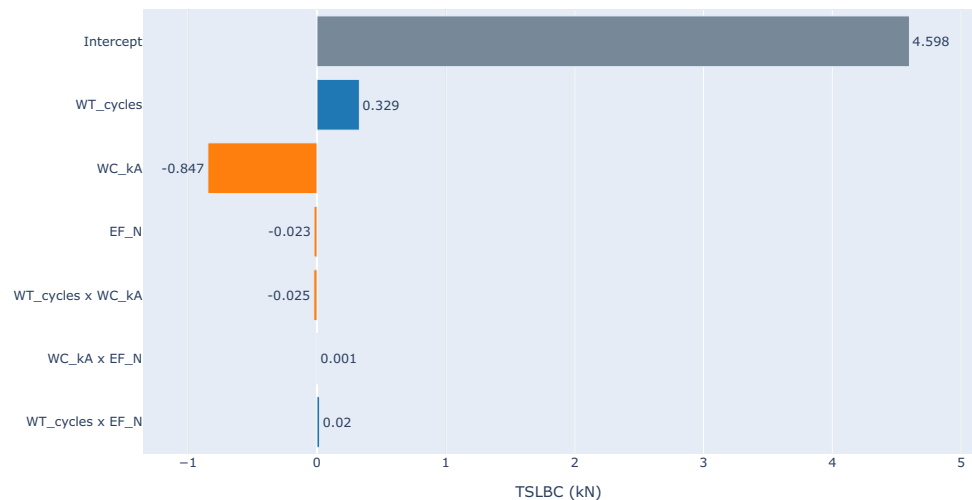


Fig. 10 Bar chart (fault diagnosis plot) showing the contributions of each variable to the predicted TSLBC of a RSW joint obtained with WT=0.06 s, WC=5.51 kA, and EF=1500 N. The experimental TSLBC is 5.500 kN, while the predicted TSLBC using the

EBM model is 5.354 kN. The prediction is obtained as the sum of all variable contributions, i.e., $5.354 = 4.598 - 0.567 + 1.375 - 0.023 - 0.065 + 0.046 - 0.01$

Fig. 11 Bar chart showing the contributions of each variable to the predicted TSLBC of a RSW joint obtained with WT=0.18 s, WC=2.88 kA, and EF=1500 N. The experimental TSLBC is 4.490 kN, while the predicted TSLBC using the EBM model is 4.053 kN. Note that the prediction $4.053 \text{ kN} = 4.598 + 0.329 - 0.847 - 0.023 - 0.025 + 0.001 + 0.02$



it may be time to start considering the EBM as the reference algorithm for the TSLBC-RSW problem, the one to be chosen by default in the absence of further requirements. Besides, the use of this type of model may be potentially useful in other alternative ways (i) to capture patterns and bring to light new phenomena and (ii) to formulate hypothesis/theoretical models when the latter do not exist.

5 Conclusions

In this work, it is shown that the explainable boosting machine—an interpretable algorithm—provides statistically as good or even better results for both regression and classification than

the best-performing black-box algorithms proposed so far in the scientific literature for the RSW quality prediction problem. In particular:

- for the regression problem of predicting the TSLBC from the key welding parameters WT, WC, and EF, the regression EBM is the highest-performing algorithm (mean MSE = 0.031, std = 0.013), followed by the regression GBM (mean MSE = 0.033, std = 0.015) and the regression RF (mean MSE = 0.034, std = 0.017), and
- for the classification of RSW joints as valid or invalid from their welding parameters WT, WC, and EF, the best-performing algorithm is the deep learning classifier (mean classification error = 0.0658, std = 0.0393),

followed by the AdaBoost classifier (mean classification error = 0.0658, std = 0.0584) and the classification EBM (mean classification error = 0.0697, std = 0.0539).

These results were analyzed statistically from both frequentist and Bayesian perspectives. No statistically significant differences were found either between the regression EBM and the second best-performing regressor, or between the classification EBM and the best-performing classifier with either approach.

Importantly, in addition to providing high levels of accuracy, EBMs have the additional advantage of being interpretable, both at the level of the overall behavior of the model and at the level of individual predictions:

- From the overall behavior perspective, in this work, the analysis of the relative influence of the welding variables on the TSLBC has indicated that WC and WT are the most relevant (WC being 3.7 times more important than WT). Besides, the analysis of interaction effects has served to explore the nonlinear nature of the TSLBC response—related to saturation effects due to expulsion and/or quadric effects with respect to WC.
- At the level of individual prediction—i.e., when analyzing the quality of a given RSW joint—the EBM allows to univocally determine how the value of each welding parameter contributed to the response obtained, thus being very useful for fault diagnosis purposes.

In a nutshell, the results of this work demonstrate that the EBM, due to its highest accuracy performance and high interpretability, is one of the most valuable approaches that can be currently used as a decision support tool for establishing RSW operating parameters, engineering failure analysis, and manufacturing quality control purposes.

Author contribution José I. Santos: writing—original draft preparation; writing—review and editing; computational and statistical design; software implementation

Oscar Martín: writing—original draft preparation; writing—review and editing; experimental design; investigation (experimental testing); conceptualization

Virginia Ahedo: writing—original draft preparation; writing—review and editing; computational and statistical design.

Pilar De Tiedra: investigation (experimental testing)

José M. Galán: writing—original draft preparation; writing—review and editing; computational and statistical design; conceptualization; supervision

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. The authors acknowledge financial support from the Spanish Ministry of Science, Innovation and Universities (Excellence Network RED2018-102518-T), the Spanish State Research Agency (PID2020-118906 GB-I00/AEI/<https://doi.org/10.13039/501100011033>), and the Fundación Bancaria Caixa D. Estalvis

I Pensions de Barcelona, La Caixa (2020/00062/001). In addition, we acknowledge support from the Santander Supercomputación group (University of Cantabria) that 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.

Availability of data, material and code The data that support the findings of this study is part of an ongoing project and are available from Oscar Martín, upon request.

Declarations

Ethics approval The authors declare that this manuscript was not submitted to more than one journal for simultaneous consideration. Also, the submitted work is original and has not been published elsewhere in any form or language.

Consent to participate and consent for publication The authors declare that they participated in this paper willingly and the authors declare to consent to the publication of this paper.

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Becker N, Gilgert J, Petit EJ, Azari Z (2014) The effect of galvanizing on the mechanical resistance and fatigue toughness of a spot welded assembly made of AISI410 martensite. *Mater Sci Eng A* 596:145–156. <https://doi.org/10.1016/j.msea.2013.12.008>
2. Soomro IA, Pedapati SR, Awang M (2021) Optimization of post-weld tempering pulse parameters for maximum load bearing and failure energy absorption in dual phase (DP590) steel resistance spot welds. *Mater Sci Eng, A* 803:140713. <https://doi.org/10.1016/j.msea.2020.140713>
3. Xia Y-J, Su Z-W, Li Y-B et al (2019) Online quantitative evaluation of expulsion in resistance spot welding. *J Manuf Process* 46:34–43. <https://doi.org/10.1016/j.jmapro.2019.08.004>
4. Soomro IA, Pedapati SR, Awang M (2022) A review of advances in resistance spot welding of automotive sheet steels: emerging methods to improve joint mechanical performance. *Int J Adv Manuf Technol* 118:1335–1366. <https://doi.org/10.1007/s00170-021-08002-5>
5. Janardhan G, Kishore K, Dutta K, Mukhopadhyay G (2020) Tensile and fatigue behavior of resistance spot-welded HSLA steel sheets: effect of pre-strain in association with dislocation density. *Mater Sci Eng A* 793:139796. <https://doi.org/10.1016/j.msea.2020.139796>
6. Ertek Emre H, Bozkurt B (2020) Effect of Cr-Ni coated Cu-Cr-Zr electrodes on the mechanical properties and failure modes of

- TRIP800 spot weldments. *Eng Fail Anal* 110:104439. <https://doi.org/10.1016/j.engfailanal.2020.104439>
7. Sun X, Zhang Q, Wang S et al (2020) Effect of adhesive sealant on resistance spot welding of 301L stainless steel. *J Manuf Process* 51:62–72. <https://doi.org/10.1016/j.jmapro.2020.01.033>
 8. Qi L, Li F, Zhang Q et al (2021) Improvement of single-sided resistance spot welding of austenitic stainless steel using radial magnetic field. *J Manuf Sci Eng* 143:031004. <https://doi.org/10.1115/1.4048048>
 9. Wen J, de Jia H, Wang CS (2019) Quality estimation system for resistance spot welding of stainless steel. *ISIJ Int* 59:2073–2076. <https://doi.org/10.2355/isijinternational.ISIJINT-2019-002>
 10. Özyürek D (2008) An effect of weld current and weld atmosphere on the resistance spot weldability of 304L austenitic stainless steel. *Mater Des* 29:597–603. <https://doi.org/10.1016/j.matdes.2007.03.008>
 11. Zhou M, Hu S, Zhang H (1999) Critical specimen sizes for tensile-shear testing of steel sheets. *Weld J* 78:305S–313S
 12. Bemani M, Pouranvari M (2020) Microstructure and mechanical properties of dissimilar nickel-based superalloys resistance spot welds. *Mater Sci Eng A* 773:138825. <https://doi.org/10.1016/j.msea.2019.138825>
 13. Martín Ó, Ahedo V, Santos JI et al (2016) Quality assessment of resistance spot welding joints of AISI 304 stainless steel based on elastic nets. *Mater Sci Eng A* 676:173–181. <https://doi.org/10.1016/j.msea.2016.08.112>
 14. Martín Ó, De Tiedra P, San-Juan M (2017) Combined effect of resistance spot welding and precipitation hardening on tensile shear load bearing capacity of A286 superalloy. *Mater Sci Eng A* 688:309–314. <https://doi.org/10.1016/j.msea.2017.02.015>
 15. Oliveira JP, Ponder K, Brizes E et al (2019) Combining resistance spot welding and friction element welding for dissimilar joining of aluminum to high strength steels. *J Mater Process Technol* 273:116192. <https://doi.org/10.1016/j.jmatprotec.2019.04.018>
 16. Hasanbaşoğlu A, Kaçar R (2007) Resistance spot weldability of dissimilar materials (AISI 316L–DIN EN 10130–99 steels). *Mater Des* 28:1794–1800. <https://doi.org/10.1016/j.matdes.2006.05.013>
 17. Kong JP, Han TK, Chin KG et al (2014) Effect of boron content and welding current on the mechanical properties of electrical resistance spot welds in complex-phase steels. *Mater Des* 54:598–609. <https://doi.org/10.1016/j.matdes.2013.08.098>
 18. Badkoobeh F, Nouri A, Hassannejad H, Mostaan H (2020) Microstructure and mechanical properties of resistance spot welded dual-phase steels with various silicon contents. *Mater Sci Eng A* 790:139703. <https://doi.org/10.1016/j.msea.2020.139703>
 19. Zhou K, Yao P (2019) Overview of recent advances of process analysis and quality control in resistance spot welding. *Mech Syst Signal Process* 124:170–198. <https://doi.org/10.1016/j.ymssp.2019.01.041>
 20. Pereda M, Santos JI, Martín Ó, Galán JM (2015) Direct quality prediction in resistance spot welding process: sensitivity, specificity and predictive accuracy comparative analysis. *Sci Technol Weld Join* 20:679–685. <https://doi.org/10.1179/1362171815Y.0000000052>
 21. Martín O, Ahedo V, Santos JI, Galán JM (2022) Comparative study of classification algorithms for quality assessment of resistance spot welding joints from pre and post-welding inputs. *IEEE Access* 10:6518–6527. <https://doi.org/10.1109/ACCESS.2022.3142515>
 22. Tercan H, Meisen T (2022) Machine learning and deep learning based predictive quality in manufacturing: a systematic review. *J Intell Manuf*. <https://doi.org/10.1007/s10845-022-01963-8>
 23. Zamanzad Gavidel S, Lu S, Rickli JL (2019) Performance analysis and comparison of machine learning algorithms for predicting nugget width of resistance spot welding joints. *Int J Adv Manuf Technol* 105:3779–3796. <https://doi.org/10.1007/s00170-019-03821-z>
 24. Adadi A, Berrada M (2018) Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access* 6:52138–52160. <https://doi.org/10.1109/ACCESS.2018.2870052>
 25. Borrego-Díaz J, Galán Páez J (2022) Knowledge representation for explainable artificial intelligence. *Complex Intell Syst* 8:1579–1601. <https://doi.org/10.1007/s40747-021-00613-5>
 26. Rudin C, Radin J (2019) Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition. *Harv Data Sci Rev* 1. <https://doi.org/10.1162/99608f92.5a8a3a3d>
 27. Martín Ó, De TP, López M et al (2009) Quality prediction of resistance spot welding joints of 304 austenitic stainless steel. *Mater Des* 30:68–77. <https://doi.org/10.1016/j.matdes.2008.04.050>
 28. Mousavi Anijdan SH, Sabzi M, Ghobeiti-Hasab M, Roshan-Ghiyas A (2018) Optimization of spot welding process parameters in dissimilar joint of dual phase steel DP600 and AISI 304 stainless steel to achieve the highest level of shear-tensile strength. *Mater Sci Eng A* 726:120–125. <https://doi.org/10.1016/j.msea.2018.04.072>
 29. Wang X, Guan S, Hua L et al (2019) Classification of spot-welded joint strength using ultrasonic signal time-frequency features and PSO-SVM method. *Ultrasonics* 91:161–169. <https://doi.org/10.1016/j.ultras.2018.08.014>
 30. Valaee-Tale M, Sheikhi M, Mazaheri Y et al (2020) Criterion for predicting expulsion in resistance spot welding of steel sheets. *J Mater Process Technol* 275:116329. <https://doi.org/10.1016/j.jmatprotec.2019.116329>
 31. Chen T, Ling Z, Wang M, Kong L (2020) Effect of a slightly concave electrode on resistance spot welding of Q&P1180 steel. *J Mater Process Technol* 285:116797. <https://doi.org/10.1016/j.jmatprotec.2020.116797>
 32. Dejans A, Kurtov O, Van Rymenant P (2021) Acoustic emission as a tool for prediction of nugget diameter in resistance spot welding. *J Manuf Process* 62:7–17. <https://doi.org/10.1016/j.jmapro.2020.12.002>
 33. Deng L, Li Y, Cai W et al (2020) Simulating thermoelectric effect and its impact on asymmetric weld nugget growth in aluminum resistance spot welding. *J Manuf Sci Eng* 142:091001. <https://doi.org/10.1115/1.4047243>
 34. Islam MR, Ahmed MU, Barua S, Begum S (2022) A systematic review of explainable artificial intelligence in terms of different application domains and tasks. *Appl Sci (Switzerland)* 12. <https://doi.org/10.3390/app12031353>
 35. Roscher R, Bohn B, Duarte MF, Garcke J (2020) Explainable machine learning for scientific insights and discoveries. *IEEE Access* 8:42200–42216. <https://doi.org/10.1109/ACCESS.2020.2976199>
 36. Nori H, Jenkins S, Koch P, Caruana R (2019) InterpretML: a unified framework for machine learning interpretability. 1–8
 37. Ghanbari HR, Shariati M, Sanati E, Masoudi Nejad R (2022) Effects of spot welded parameters on fatigue behavior of ferrite-martensite dual-phase steel and hybrid joints. *Eng Fail Anal* 134:106079. <https://doi.org/10.1016/j.engfailanal.2022.106079>
 38. ISO 14273 (2000) Specimen dimensions and procedure for shear testing resistance spot, seam and embossed projection welds
 39. ASTM A 262–91 (1993) Standard practices for detecting susceptibility to intergranular attack in austenitic stainless steels
 40. Marashi P, Pouranvari M, Amirabdollahian S et al (2008) Microstructure and failure behavior of dissimilar resistance spot welds between low carbon galvanized and austenitic stainless steels. *Mater Sci Eng A* 480:175–180. <https://doi.org/10.1016/j.msea.2007.07.007>

41. Martín Ó, de Tiedra P, San-Juan M (2019) Effect of Widmanstätten η phase on tensile shear strength of resistance spot welding joints of A286 superalloy. *Metall Res Technol* 116:302. <https://doi.org/10.1051/metal/2018095>
42. JIS Z 3140 (1989) Method of inspection for spot weld
43. Mansour T (1991) Ultrasonic testing of spot welds in thin gage steel. In: McIntire P (ed) *Nondestructive Testing Handbook*. Vol. 7: Ultrasonic Testing. American Society for Nondestructive Testing, Metals Park, pp 557–568
44. Martín Ó, Pereda M, Santos JI, Galán JM (2014) Assessment of resistance spot welding quality based on ultrasonic testing and tree-based techniques. *J Mater Process Technol* 214:2478–2487. <https://doi.org/10.1016/j.jmatprotec.2014.05.021>
45. Hastie TJ, Tibshirani RJ (1990) *Generalized additive models*. Chapman & Hall/CRC Monographs on statistics and Applied Probability, New York
46. Lou Y, Caruana R, Gehrke J, Hooker G (2013) Accurate intelligible models with pairwise interactions. *Proc ACM SIGKDD Int Conf Knowl Discov Data Min Part F1288:623–631*. <https://doi.org/10.1145/2487575.2487579>
47. Caruana R, Lou Y, Gehrke J et al (2015) Intelligible models for HealthCare. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, USA, pp 1721–1730
48. Magunia H, Lederer S, Verbuecheln R et al (2021) Machine learning identifies ICU outcome predictors in a multicenter COVID-19 cohort. *Crit Care* 25:1–14. <https://doi.org/10.1186/s13054-021-03720-4>
49. Morgan HE, Wang K, Dohopolski M et al (2021) Exploratory ensemble interpretable model for predicting local failure in head and neck cancer: the additive benefit of CT and intra-treatment cone-beam computed tomography features. *Quant Imaging Med Surg* 11:4781–4796. <https://doi.org/10.21037/qims-21-274>
50. Sarica A, Quattrone A, Quattrone A (2021) Explainable boosting machine for predicting Alzheimer's disease from MRI hippocampal subfields. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer International Publishing, pp 341–350
51. Wang H, Huang Z, Zhang D et al (2020) Integrating co-clustering and interpretable machine learning for the prediction of intravenous immunoglobulin resistance in Kawasaki disease. *IEEE Access* 8:97064–97071. <https://doi.org/10.1109/ACCESS.2020.2996302>
52. Maxwell AE, Sharma M, Donaldson KA (2021) Explainable boosting machines for slope failure spatial predictive modeling. *Remote Sens (Basel)* 13:4991. <https://doi.org/10.3390/rs13244991>
53. Lou Y, Caruana R, Gehrke J (2012) Intelligible models for classification and regression. In: *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '12*. ACM Press, New York, New York, USA, p 150
54. Guo P, Zhu Q, Kang J et al (2022) Quality assessment of RSW based on transfer learning and imbalanced multi-class classification algorithm. *IEEE Access* 1–1. <https://doi.org/10.1109/ACCESS.2022.3212410>
55. Anderssen E, Dyrstad K, Westad F, Martens H (2006) Reducing over-optimism in variable selection by cross-model validation. *Chemom Intell Lab Syst* 84:69–74. <https://doi.org/10.1016/j.chemolab.2006.04.021>
56. Varma S, Simon R (2006) Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* 7:91. <https://doi.org/10.1186/1471-2105-7-91>
57. Gómez-Ríos A, Luengo J, Herrera F (2017) A study on the noise label influence in boosting algorithms: Adaboost, GBM and XGBoost. *Lecture notes in computer science* 10334 LNCS:268–280. https://doi.org/10.1007/978-3-319-59650-1_23
58. Freund Y, Schapire RE (1996) Experiments with a new boosting algorithm. *Proceedings of the 13th International Conference on Machine Learning* 148–156. <https://doi.org/10.5555/3091696.3091715>
59. Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Ann Stat* 29:1189–1232
60. Chen T, Guestrin C (2016) XGBoost. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, USA, pp 785–794
61. Gao X, Lin C (2021) Prediction model of the failure mode of beam-column joints using machine learning methods. *Eng Fail Anal* 120:105072. <https://doi.org/10.1016/j.engfailanal.2020.105072>
62. Feng D-C, Wang W-J, Mangalathu S, Taciroglu E (2021) Interpretable XGBoost-SHAP machine-learning model for shear strength prediction of squat RC walls. *J Struct Eng* 147. [https://doi.org/10.1061/\(ASCE\)ST.1943-541X.0003115](https://doi.org/10.1061/(ASCE)ST.1943-541X.0003115)
63. Qiu Y, Zhou J, Khandelwal M et al (2021) Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration. *Eng Comput*. <https://doi.org/10.1007/s00366-021-01393-9>
64. Chen K, Chen H, Liu L, Chen S (2019) Prediction of weld bead geometry of MAG welding based on XGBoost algorithm. *Int J Adv Manuf Technol* 101:2283–2295. <https://doi.org/10.1007/s00170-018-3083-6>
65. Zhang Z, Huang Y, Qin R et al (2021) XGBoost-based on-line prediction of seam tensile strength for Al-Li alloy in laser welding: experiment study and modelling. *J Manuf Process* 64:30–44. <https://doi.org/10.1016/j.jmapro.2020.12.004>
66. Breiman L (2001) Random forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
67. Breiman L (1996) Bagging predictors. *Mach Learn* 24:123–140. <https://doi.org/10.1007/BF00058655>
68. Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers. In: *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*. ACM Press, pp 144–152
69. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444. <https://doi.org/10.1038/nature14539>
70. Howard J, Gugger S (2020) Fastai: a layered API for deep learning. *Information* 11:108. <https://doi.org/10.3390/info11020108>
71. Le Q, Ngiam J, Coates A, et al (2011) On optimization methods for deep learning. In: Getoor L, Scheffer T (eds) *ICML'11: Proceedings of the 28th International Conference on International Conference on Machine Learning*. Omnipress, Madison, WI, USA, pp 265–272
72. Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol* 67:301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>
73. Carrasco J, García S, Rueda MM et al (2020) Recent trends in the use of statistical tests for comparing swarm and evolutionary computing algorithms: practical guidelines and a critical review. *Swarm Evol Comput* 54. <https://doi.org/10.1016/j.swevo.2020.100665>
74. Nadeau C, Bengio Y (2003) Inference for the generalization error. *Mach Learn* 52:239–281. <https://doi.org/10.1023/A:1024068626366>
75. Benavoli A, Corani G, Demšar J, Zaffalon M (2017) Time for a change: a tutorial for comparing multiple classifiers through Bayesian analysis. *J Mach Learn Res* 18:1–36
76. Pashazadeh H, Gheisari Y, Hamed M (2016) Statistical modeling and optimization of resistance spot welding process parameters using neural networks and multi-objective genetic algorithm. *J Intell Manuf* 27:549–559. <https://doi.org/10.1007/s10845-014-0891-x>
77. Zhao D, Ivanov M, Wang Y et al (2021) Multi-objective optimization of the resistance spot welding process using a hybrid

- approach. *J Intell Manuf* 32:2219–2234. <https://doi.org/10.1007/s10845-020-01638-2>
78. Aslanlar S, Ogur A, Ozsarac U et al (2007) Effect of welding current on mechanical properties of galvanized chromided steel sheets in electrical resistance spot welding. *Mater Des* 28:2–7. <https://doi.org/10.1016/j.matdes.2005.06.022>
79. Aslanlar S, Ogur A, Ozsarac U, Ilhan E (2008) Welding time effect on mechanical properties of automotive sheets in electrical resistance spot welding. *Mater Des* 29:1427–1431. <https://doi.org/10.1016/j.matdes.2007.09.004>
80. Kong JP, Kang CY (2016) Effect of alloying elements on expulsion in electric resistance spot welding of advanced high strength steels. *Sci Technol Weld Joining* 21:32–42. <https://doi.org/10.1179/1362171815Y.0000000057>
81. Xing B, Xiao Y, Qin QH (2018) Characteristics of shunting effect in resistance spot welding in mild steel based on electrode displacement. *Measurement* 115:233–242. <https://doi.org/10.1016/j.measurement.2017.10.049>
82. Xing B, Xiao Y, Qin QH, Cui H (2018) Quality assessment of resistance spot welding process based on dynamic resistance signal and random forest based. *Int J Adv Manuf Technol* 94:327–339. <https://doi.org/10.1007/s00170-017-0889-6>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.