IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 20 September 2022, accepted 3 October 2022, date of publication 5 October 2022, date of current version 4 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3212410

RESEARCH ARTICLE

Quality Assessment of RSW Based on Transfer Learning and Imbalanced Multi-Class Classification Algorithm

PEIJIN GUO[1], QINMIAO ZHU[2,3], JINGRAN KANG[4], YUHUI WANG[2,3], AND WENQIANG HU[1]

[1]School of Software, Huazhong University of Science and Technology, Wuhan 430074, China
[2]School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
[3]Wuhan Digital Design and Manufacturing Innovation Center Co., Ltd., Wuhan 430206, China
[4]Geely Automobile Research Institute (Ningbo) Company Ltd., Ningbo 315000, China

Corresponding author: Wenqiang Hu (hwq_hust@mail.hust.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFB1710700.

ABSTRACT In automobile manufacturing, the quality assessment of resistance spot welding (RSW) plays a decisive role in the quality and safety of products. Recently, it has become very popular to use machine learning to evaluate the quality of welding nuggets. However, there are two obstacles: data imbalance caused by limited defective samples, and data shortage due to expensive time and labor costs. This paper proposes a novel method. On one hand, the self-paced ensemble (SPE) algorithm for binary classification is improved to handle imbalanced multi-class classification of quality levels. On the other hand, an instance-based ensemble transfer learning approach is proposed to predict the tensile-shear strength of RSW for precise control of the weld quality. In detail, a quality level identification model is formulated with the process and material parameters as the input at first. Secondly, an explainable algorithm SHapley Additive exPlanations (SHAP) was introduced to anatomize the impacts of welding parameters on the welding quality predictions. Finally, a hybrid dataset including actual historic production data and 454 spot-welding cases is constructed, and then the eXtreme Gradient Boosting (XGBoost) is introduced as the base learner of TrAdaBoost.R2 to train the prediction model. Compared with conventional methods, the SPE provides the greatest macro geometric-mean score of **0.923**, and the proposed regression model yields superior accuracy R^2 of **0.952**, which shows the potential of assisting welding process design.

INDEX TERMS Resistance spot welding, class-imbalanced classification, transfer learning.

I. INTRODUCTION

Resistance spot welding(RSW) is a material connection technology widely used in the automotive industry. The welding quality is very important to the overall safety of the body-in-white, which directly affects the reliability, performance, and manufacturing cost of the vehicle, and is closely related to the owner's safety [1]. RSW process for automotive welding, traditional manual quality inspection of welding nuggets can not cover all welding points, so it is necessary to increase the spot welding density on the car body to make up for potential safety hazards, which leads to low production efficiency and high error rates. Meanwhile, facing the requirements of lightweight car bodies [2], more and more new material combinations are applied to car design, and correspondingly more weldability tests are needed, resulting in increased production costs. The process parameters obtained by the destructive test are very reliable, but the empirical formula(e.g., response surface methodology [3]) cannot be directly transferred to the new welding design. Therefore, it is of interest to the resistance spot welding industry to reduce the number of tests on the premise of ensuring quality.

To deal with the above difficulties, numerical simulation and data-driven modeling are two widely used methods. The finite element analysis model based on the knowledge of welding experts can simulate the resistance spot

The associate editor coordinating the review of this manuscript and approving it for publication was Alberto Cano.

welding process, but the numerical simulation is obstructed by multiple physical factors of the industrial production environment [4]. The data-driven model is independent of the physical model, and it is suitable for the research object with complex structure, so it has attracted extensive attention. The applied data-driven methods mainly include two types: image recognition and monitoring the signals or parameters of the welding machine. The first type is spot welding quality inspection based on machine vision. Xiao et al. [5] take the welding spot image as the input of a convolutional neural network(CNN) to identify the appearance of the nugget. Dai et al. [6] mark different quality spot welds on the body-in-white according to the size of the nugget diameter, to detect the quality of spot welding based on the You Only Look Once(YOLO) algorithm. However, the training of deep neural networks depends on a large number of labeled images.

The second type is to collect physical signals in the welding process, to obtain the key information that affects the welding quality. Or take welding process parameters as input features. Dai et al. [7] use the preprocessed dynamic resistance signals to train the convolutional neural network to achieve quality assessment of spot welding. Zhou et al. [8] use principal component analysis to reduce the dimensionality of the original dynamic resistance signals to obtain input features, and use machine learning algorithms to train the model for spatter prediction during resistance spot welding. Dejans et al. [9] use the acoustic signal waveform collected from the welding process, the frequency components related to the welding physical process are extracted from it, and these amplitudes at a specific frequency are used to further build a prediction model of the nugget diameter of resistance spot welding. Although the nugget diameter can be predicted nondestructively, the prediction accuracy of this method is lower than that based on machine learning methods. Ghafarallahi et al. [10] compare the performance of the three methods, i.e., Artificial Neural Network (ANN), Multilayer Perceptron (MLP), and mathematical formula calculation in diameter prediction of three-sheet spot welds, among which ANN achieves the highest accuracy. Kim et al. [11] propose that the thickness, tensile strength, and yield strength of the material should be combined with the features extracted from the electrode displacement curve and dynamic resistance curve, and a polynomial relationship should be established with the tensile shear strength and indentation depth, respectively, and then the logistic regression algorithm can be used to predict the weld quality. To eliminate redundant information in data, Dang et al. [12] propose a smart framework based on machine learning algorithms. Collected from different welding environments, the inconsistency of the characteristics and distribution of the data does not affect the performance of the model.

Although the data-driven modeling can achieve satisfactory results, there are still some limitations:

1) **Imbalance of data.** With continuous production, the factory will quickly accumulate qualified welding data, but the proportion of defective data and unqualified data is still low, resulting in an imbalanced data set.
2) **Shortage of data.** It is time-consuming and difficult to construct a large annotation data set, and environmental noise results in different responses to the same welding process design parameters.

On one hand, to avoid data imbalance, previous works [13], [14] tend to use data with relatively balanced classes to train models. After some time, as more and more welding failure data are discarded, effective information in welding quality prediction may be lost. However, traditional machine learning-based methods (such as support vector machine) have difficulty distinguishing minority classes when the dataset classes are imbalanced [15]. To overcome the problem, this article introduces the Self-paced ensemble (SPE) [16] method for the imbalanced multi-class classification prediction of nugget quality in the RSW process. A classification algorithm based on SPE can improve the recognition rate of rare minority classes such as Pseudo soldering. These minority classes need to be accurately identified to avoid failed welds in the future.

On the other hand, nugget diameter, an important indicator to measure the quality of the weld, can be assessed using non-destructive testing techniques after welding. However, when the size and distribution of the defects inside and outside the nugget are different, the nugget diameter can't accurately represent the nugget quality. Tensile shear strength is a robust indicator, but requires destructive tear testing to measure. To reduce the amount of training data and improve the generalization ability of the model, many efforts have been made to develop a transfer learning prediction model. Pan et al. [17] show that transfer learning only needs a certain relationship between the source domain and the target domain, so the knowledge and features learned in the source domain can help train the model in the target domain, to realize the knowledge transfer between different domains. This brings new inspiration to RSW quality assessment. To solve these problems, a transfer learning (TR) method based on TrAdaBoost.R2 [18] and XGBoost is proposed to predict the tensile shear strength of nuggets in RSW. By making effective use of historical data, the TR solves the modeling problem of different welding material combinations and enhances the generalization ability of the model. Through the TR, the prediction ability of RSW quality is improved to cope with different production environments and few-shot datasets, thus ensuring the accuracy of the prediction. The contributions of this paper are summarized as follows:

1) A novel imbalanced multi-class classification algorithm, self-paced ensemble (SPE), is proposed to identify the quality level of nuggets in resistance spot welding (RSW). During the resampling process, a self-paced factor is introduced to optimize the sampling strategy considering the classification hardness distribution on the dataset. Experimental results show that the SPE achieves the highest macro $G - mean$ of 0.923 and macro-average AUC of 0.976. Welding

process parameters and material parameters are input into the model, which is convenient for adjusting these key parameters in the future, realizing accurate welding quality control, and reducing welding faults.

2) For the assessment of tensile shear strength of nuggets in RSW, an instance-based algorithm TrAdaBoost.R2 is introduced. Compared with existing data-driven methods, TrAdaBoost.R2 achieves better evaluation performance. It achieves the highest R^2 value of 0.952 and the lowest *RMSE* fitting error of 2.692 *KN* using a small amount of data in the target domain. From the application point of view, the proposed method helps to reduce the number of weldability tests, assists the design of welding process parameters, and reduces the cost of labeling data.

The rest organization of this paper is as follows: Section II introduces the experimental data and the proposed methods. In Section III, the performance of the proposed methods is verified. Section IV is the discussion and future research. Section V is the conclusion of this paper.

II. METHODOLOGY

The purpose of this study is to effectively evaluate the quality of RSW. To achieve the study objective, this research is divided into three parts, including data preprocessing, welding quality levels identification based on an imbalanced multi-class classification algorithm, and tensile shear strength prediction based on transfer learning.

The overall analysis framework is shown in Figure 1. First, the original data collected from the RSW process and the data collected from the public database are preprocessed. Then the SPE is combined with SHAP [19] to build an interpretable model, and SHAP is used to further explain the importance and impact of input features on model predictions. A consistent feature dimension in the source domain and the target domain is constructed manually. The labels of the source and target domains are nugget diameter and tensile shear strength, respectively. Finally, a tensile shear strength prediction model is constructed based on TrAdaBoost.R2 to assess the quality of nuggets.

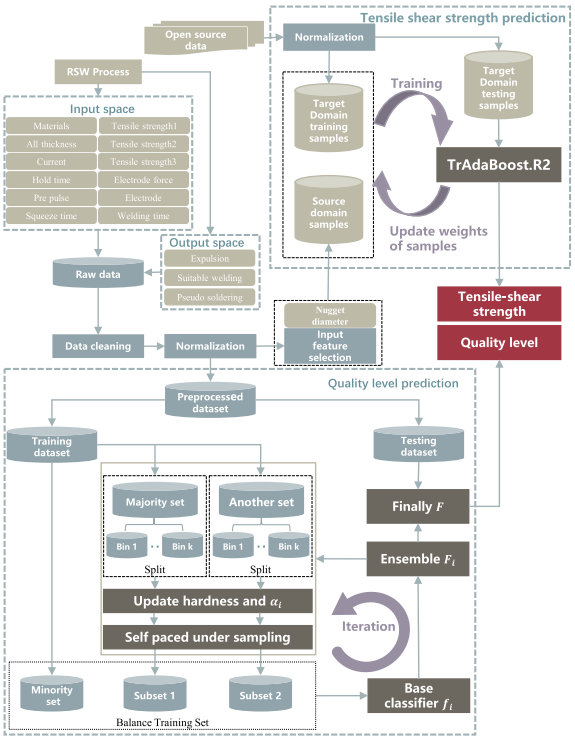

FIGURE 1. Workflow of the proposed research methodology.

A. DATA DESCRIPTION AND PREPROCESSING

In this paper, two groups of data are constructed. The first dataset 1 consists of the historical experiment welding data samples, while the second dataset 2 is collected from existing literature [3], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37]. Dataset 1 contains a large number of different models of material, the welding materials are mainly hot-dip galvanized steel (e.g., DC56+D et al.). Dataset 1 contains 4848 RSW samples (Table 1.), including 4362 suitable welding samples, 204 pseudo soldering samples, and 282 welding expulsion samples. Each sample corresponds to a nugget diameter. The process parameters in dataset 1 include the number of

TABLE 1. Welding parameters of dataset 1.

Parameters	Feture with units
Material	Number of materials
	Total thickness (mm)
	Tensile strength1 (MPa)
	Tensile strength2 (MPa)
	Tensile strength3 (MPa)
Welding process	Welding time ($msec$)
	Current (kA)
	Electrode force (kN)
	Electrode tip (mm)
	Pre pulse ($msec$)
	Squeeze time ($msec$)
	Hold time ($msec$)
Welding quality	Suitable welding
	Expulsion
	Pseudo soldering
	Nugget diameter (cm)

TABLE 2. Welding parameters of dataset 2.

Parameters	Feture with units
Material	Total thickness (mm)
Welding process	Welding time ($msec$)
	Current (kA)
	Electrode force (kN)
Welding quality	Tensile shear strength (kN)

welding materials, total thickness of all welding materials, tensile properties of welding materials, welding time, current, electrode force, electrode diameter, pre-pulse, squeeze time, and hold time.

Dataset 2 contains a series of metals including stainless steel [3], [20], [21], [22], [23], [24], [25], low carbon steel [20], [26], [27], [28], [29], [30], [31], [32], galvanized steel sheet [33], [34], [35], high strength steel [36], [37]. Dataset 2 contains 454 two-layer sheet RSW samples (Table 2.), these samples were obtained by resistance spot welding of two layers of sheet metal. The input feature parameters in dataset 2 include the total thickness of all welding materials, welding time, current, electrode force, electrode diameter, hold time, and the output tensile shear strength.

Due to defects or contamination on the surface of the weldment, the electrode is bonded to the material during the welding process. This type of data in dataset 1 is filtered out. In order to eliminate the influence of dimension, all features in dataset 1 and dataset 2 are transformed with standard Min-Max normalization.

B. IMPROVED IMBALANCED MULTI-CLASS CLASSIFICATION ALGORITHM BASED SELF-PACED ENSEMBLE

Classical machine learning algorithms assume that the numbers of samples in different categories are similar, so when the number of samples in different categories in the data set is imbalanced, these algorithms will give priority to a class with more samples. In the manufacturing process, defective welding nuggets should be avoided, so we pay more attention to the ability of the model to correctly distinguish samples belonging to a few categories.

Cost-sensitive methods and resampling techniques are designed to deal with class imbalance problems. Cost-sensitive methods need to incorporate domain-specific expert knowledge, which is difficult to achieve in many tasks [38]. Resampling techniques include oversampling, undersampling, and a hybrid of both. These methods seldom take into account the influence of data distribution on classification performance. It is necessary to pay full attention to the sample distribution of all classes when the minority class and the majority class are equally important.

SPE imbalanced classification algorithm introduces the concept of ''classification hardness'' to portray the difficulty for a trained classifier to classify a specific sample. For samples belonging to the majority class, SPE iteratively adjusts the hardness value of the samples to mark samples with different classification difficulties. Then, the distribution of hardness values contains additional information, such as possible noise in samples with too high hardness values. Suppose F is a trained classifier, $F(x_i)$ is used to denote the classifier's output probability of x_i. We use the symbol H to represent the classification hardness function, where H is the Mean Absolute Error, which is less sensitive to outliers than the Mean Squared Error. The hardness function is given as:

$$H(x, y, F) = \frac{1}{n} \sum_{i=1}^{n} |f_i(x_i) - y_i| \tag{1}$$

The distribution of hardness H reflects the fitting of F to data set (x, y).

According to hardness values, data samples are divided into three categories: trivial samples, noise samples, and borderline samples. SPE holds an under-sampling mechanism to reduce the influence of trivial samples and noisy samples, but to expand the importance of boundary samples. For this reason, the majority samples are split into k bins, and then we resample the majority samples into a balanced subset, each bin keeping the same hardness. In order to improve the diversity of the classifier and prevent the model from overfitting, we introduce the self-paced factor α:

$$\alpha = \tan \frac{i\pi}{2n} \tag{2}$$

where, i represents the current iteration, n is the total number of iteration and set $\alpha = 0$ in the first iteration.

The idea of self-paced involves simple samples, and then gradually adds hard samples to train the model [39]. In the first few iterations, the informative borderline samples are focused, as α becomes larger the model gradually focuses more on harder samples. In another word, self-paced factor α is to reduce the sampling weight of bins with too many samples.

In **Algorithm 1**, the self-paced ensemble applied to imbalanced multi-class classification is described, and the minority dataset is represented as P, the majority dataset is represented as N, and another dataset is represented as M. SPE randomly undersamples dataset M, N, generates dataset $M_0, N_0, |M_0| = |N_0| = |P|$, and then trains the base classifier f_0 on $S_0 = M_0 \cup N_0 \cup P$. Denoting B^l the $l - th$ bin, B^l is defined as:

$$B^l = \left\{ (x, y) | \frac{l-1}{k} \leq H(x, y, F) < \frac{l}{k} \right\}, \; H \in [0, 1] \tag{3}$$

C. DESIGN OF PREDICTION ALGORITHM OF TENSILE SHEAR STRENGTH BASED ON TRANSFER LEARNING

The expensive data acquisition in car manufacturing limits the predictive accuracy of data-driven algorithms. To address this shortcoming, this study employs an instances-based domain adaptation transfer learning algorithm. Transfer learning methods do not need to retrain a model for each task, and its idea is that the knowledge learned from the source domain can be used to solve problems in the target domain.

The TrAdaBoost.R2 algorithm trains a base learner in each iteration, and dynamically adjusts the weights of instances. The weights of source domain instances with poor prediction performance will be reduced, while the weights of the target domain instances will be gradually increased. The base learners trained in iterative rounds are based on the results of the previous iteration, and the final prediction output is the weighted median of base learners. Therefore, if the prediction

Algorithm 1 Self-Paced Ensemble

Inputs: Number of base classifiers n,base classifier f, dataset P, M, N, where $|P| < |M|$,$|P| < |N|$, number of bins k, hardness function H

Initialize: Train classifier f_0 using dataset S_0

for $i = 1$ to n **do**

Ensemble base classifiers $F_i(x) = \frac{1}{i}\sum_{j=0}^{i-1} f_j(x)$

Divide dataset M into k bins based on $H_M(x, y, F_i)$, divide dataset N into k bins based on $H_N(x, y, F_i)$, obtaining $B_M^1, B_M^2, \ldots, B_M^k$ and $B_N^1, B_N^2, \ldots, B_N^k$

Average hardness contribution in l-th bin:

$$h_M^l = \sum_{s \in B_M^l} H_M(x_s, y_s, F_i)/|B_M^l|,$$

$$h_N^l = \sum_{s \in B_N^l} H_N(x_s, y_s, F_i)/|B_N^l|, l = 1, \ldots, k$$

Update $\alpha = tan(i\pi/2n)$.

Unnormalized sampling weight of lth bin: $w_M^l = \frac{1}{\alpha + h_M^l}$, $w_N^l = \frac{1}{\alpha + h_N^l}$, $l = 1, \ldots, k$.

For all $l = 1 \rightarrow k$, under-sample from B_M^l with $\frac{w_M^l}{\sum_{j=1}^k w_M^j}|P|$ samples and obtains M_i, under-sample from B_N^l with $\frac{w_N^l}{\sum_{j=1}^k w_N^j}|P|$ samples and obtains N_i

Train f_i using dataset $s_i = P \cup M_i \cup N_i$

end for

Outputs: Ensemble classifier $F(x) = \frac{1}{n}\sum_{j=1}^n f_j(x)$

accuracy of the base learner is higher, the ensemble may perform better, XGBoost is selected as the base learner of TrAdaBoost.R2 to model interactions between input welding process parameters and nugget quality. XGBoost is an improved algorithm based on a gradient boosting decision tree (GBDT), which can flexibly handle various types of data such as discrete variables and continuous variables, and has strong nonlinear fitting ability and over-fitting prevention mechanisms [40].

Suppose there is a training set $T = T_s \cup T_t$, where T_s denotes the dataset 1 with the size of n and T_t denotes the dataset 2 with the size of m. In the case of containing a limited number of samples that reflect the distribution of the test set data, T_t can assist T_s in building a transfer learning model. T_s is the unknown data domain, where some samples may be beneficial to transfer and some samples may be harmful to transfer. The algorithm details of TrAdaBoost.R2 are shown in **Algorithm 2**.

D. SELECTION OF EVALUATION METRICS

From the perspective of cost and safety, incorrectly identified (false positives) and incorrectly rejected (false negatives) have different effects in different industrial application scenarios [41]. The receiver operating characteristic (ROC) and

Algorithm 2 TrAdaBoost.R2

Inputs: Base learning algorithm learner g, the maximum number of iterations N, dataset $T = T_s \cup T_t = \{(x_i, y_i)\,|i = 1, \ldots, n+m\}$,$n$ is the size of T_s, m is the size of T_t.

Initialize: Set the initial weight distribution of dataset T as: $h^1 = \left\{h_1^1, \ldots, h_i^1, \ldots, h_{n+m}^1\right\}$, $h_i^1 = \frac{1}{n+m}$,$i = 1, \ldots, n+m$.

for $i = 1$ to N **do**

Train learner $g_t(x)$ with weight distribution h^t.

Calculate the average error on T_t, if $e_t \geq 0.5$, stop and set $N = t - 1$;else $e_t = \sum_{i=n+1}^{n+m} \frac{h_i^t |y_i - g_t(x_i)|}{\sum_{i=n+1}^{n+m} h_i^t}$.

Calculate the weight coefficients of T_s and T_t. for T_s, $\beta_t = 1/\sqrt{1 + 2ln(n/N)}$, for T_t, $\beta_t = e_t/(1 - e_t)$.

Update weight distribution of dataset T:

$$h_i^{t+1} = \begin{cases} h_i^t \beta_t^{|y_i - g_t(x_i)|}, i = 1, \ldots, n \\ h_i^t \beta_t^{-|y_i - g_t(x_i)|}, i = n+1, \ldots, n+m \end{cases}$$

end for

Outputs:$G_N(x)$ is the weighted median of the last $\lceil N/2 \rceil$, using $ln(1/\beta_t)$ as the weight.

the area under the ROC curve (AUC) are chosen as evaluation metrics.

In previous studies, precision and accuracy were evaluation metrics used for classification tasks [15], [42], but due to extreme sensitivity to data distribution, precision and accuracy were not suitable for evaluating class imbalance problems. The geometric-mean score is the most reliable metric for imbalanced classification, which can reflect the comprehensive performance of models in all classes. Expulsion, Pseudo soldering, and qualified nugget are equally important for automobile manufacturing. The classification experiment uses macro $G - mean$ as the performance evaluation metric. Define n_i as the total number of samples belonging to class C_i, $mc(i, j)$ is the number of samples of class C_i judged as class C_j, and the recall and precision of class C_i can be defined:

$$P_i = \frac{mc(i, i)}{\sum_{j=1}^k mc(j, i)}$$
$$R_i = \frac{mc(i, i)}{n_i} \tag{4}$$

Then $G - mean$ is defined as:

$$G - mean = \left(\prod_{i=1}^k R_i\right)^{\frac{1}{k}} \tag{5}$$

The correlation of determination (R^2) and root mean square error ($RMSE$) are used to evaluate the predictive performance of the transfer learning model and baseline models. The formulas for these metrics are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \quad (6)$$

where y_i is the i-th observed value, $\hat{y}_i$ is the i-th predicted value, and $\bar{y}$ is the mean of all observed values.

III. EXPERIMENT AND ANALYSIS

In this section, we compared the SPE proposed in this paper with resampling, ensemble learning, and traditional classification methods to verify the performance. The SHAP is applied to identify the influence of input feature parameters on the predictions. The effectiveness of the TR model is demonstrated through a series of experiments. Python is used to implement the algorithms. The property of dataset 1 and dataset 2 are described in Table 3. and Table 4. including the minimum, maximum, mean, and standard deviation of parameters.

TABLE 3. Parameters with Min, Max, Mean and Standard deviation (Std) of dataset 1.

Feture	Min	Max	Mean	Std
Number of materials	2	3	2.528	0.499
Total thickness (mm)	1.3	6.1	2.721	0.884
Tensile strength1 (MPa)	260	1600	607.227	422.827
Tensile strength2 (MPa)	150	1600	508.325	331.035
Tensile strength3 (MPa)	0	1600	398.613	532.819
Welding time $(msec)$	130	770	392.611	119.915
Current (kA)	5.4	13.6	8.377	1.334
Electrode force (kN)	1.8	5.2	3.130	0.670
Electrode tip (mm)	16	20	18.134	1.995
Pre pulse $(msec)$	0	300	35.488	37.556
Squeeze time $(msec)$	16	800	201.420	50.290
Hold time $(msec)$	65	200	184.902	28.573
Nugget diameter (cm)	2.3	9.94	6.441	0.936

TABLE 4. Parameters with Min, Max, Mean and Standard deviation (Std) of dataset 2.

Feture	Min	Max	Mean	Std
Total thickness (mm)	1.340	4	2.275	0.783
Welding time $(msec)$	12	3000	316.096	476.265
Current (kA)	1	50	8.529	7.652
Electrode force (kN)	0.019	50	5.290	10.815
Tensile shear strength (MPa)	1.100	62.675	10.634	12.676

A. IMBALANCED CLASSIFICATION

We randomly select 30% of the complete dataset 1 as the test set, consisting of 1309 qualified nugget samples, 85 expulsion samples, and 61 pseudo-soldering samples. The remaining 70% of dataset 1 is used as the training set.

To verify the resampling performance of SPE, four resampling methods are evaluated on the training set, namely, SMOTE [43], ADASYN [44], RU (Random Under Sampling), and Clean (Neighborhood Cleaning Rule based undersampling) [45]. After resampling, to compare with previous work, we compared 10 machine learning algorithms commonly used for RSW quality assessment [8], [41], [46], [47]. Namely, K-Nearest Neighbors (KNN), support vector machine (SVM), multi-layer perceptron (MLP), Logistic Regression, decision tree, random forest, AdaBoost, GBDT, XGBoost, and LightGBM. Resample on the training set and test on the test set.

Moreover, we compared the performance of SPE with other 6 ensemble learning methods widely used in imbalance classification, namely, SMOTE Boost [48], SMOTE Bagging, Easy Ensemble [49], Balanced Random Forest [50], RUSBoost [51], and Duple-Balanced Ensemble(DUBE) [52]. We used 10-fold cross-validation on the training set and test on the test set. Default hyper-parameters are set for all models, and the number of base estimators of ensemble learning methods is uniformly set to 50 by default. To reduce randomness for all experiments in this work, we compared the average values of 100 independent experiments. The results of individual classifiers and resampling algorithms are reported in Table 5. The results of ensemble learning algorithms are reported in Table 6.

The analysis based on the experimental results is as follows:

1) As shown in Table 5., after under-sampling based on the Clean algorithm, individual classifiers perform worse, and the potential data samples which might be important to the training process are discarded. After resampling such as SMOTE, ADASYN, or RU, individual classifiers perform better. SVM performs best after SMOTE oversampling, KNN and MLP perform best after ADASYN oversampling, while other tree-based algorithms perform best after SPE undersampling. SMOTE depends on the distance among local samples, but the distance among data samples collected from welding production workshops is difficult to define. Compared with the original samples, the generated samples may still have noise, because ADASYN is sensitive to outliers.
2) As shown in Table 6., except SPE, the Balanced Random Forest is superior to resampling methods and other ensemble learning models. Easy Ensemble and Balanced Random Forest are hybrid algorithms that combine under-sampling and ensemble. Easy Ensemble only randomly selects the majority class samples, which will result in the loss of information, while Balanced Random Forest randomly selects all class samples. SMOTE Bagging and SMOTE Boost both integrate SMOTE and ensemble, while SMOTE Boost performs worse than conventional methods such as Random Forest, indicating that SMOTE has negative effects on the ensemble. RUSBoost is a hybrid approach of undersampling and ensemble learning.

TABLE 5. Multi-class classification performance (macro *G − mean* and macro-average *AUC*) of individual classifier and resampling algorithms.

Model	Original		SMOTE		ADASYN		RU		Clean		SPE	
	GM	*AUC*	*GM*	*AUC*	*GM*	*AUC*	*GM*	*AUC*	*GM*	*AUC*	*GM*	*AUC*
KNN	0.534	0.850	0.726	0.856	0.749	0.857	0.636	0.759	0.523	0.688	0.683	0.801
SVM	0.471	0.723	0.677	0.771	0.653	0.771	0.607	0.730	0.471	0.629	0.609	0.742
MLP	0.516	0.781	0.784	0.894	0.800	0.881	0.628	0.727	0.507	0.729	0.768	0.844
LogisticRegression	0.471	0.755	0.786	0.854	0.783	0.856	0.786	0.857	0.471	0.748	0.794	0.861
DecisionTree	0.783	0.861	0.818	0.881	0.820	0.880	0.859	0.862	0.689	0.696	0.915	0.973
RandomForest	0.774	0.945	0.831	0.954	0.828	0.955	0.876	0.941	0.667	0.855	0.904	0.964
AdaBoost	0.608	0.805	0.778	0.780	0.775	0.776	0.747	0.770	0.515	0.718	0.834	0.833
GBDT	0.676	0.959	0.878	0.953	0.881	0.951	0.884	0.945	0.644	0.916	0.918	0.964
LightGBM	0.788	0.968	0.839	0.965	0.837	0.965	0.889	0.942	0.690	0.903	0.918	0.969
XGBoost	0.785	0.967	0.823	0.965	0.844	0.964	0.879	0.942	0.694	0.919	**0.923**	**0.976**

TABLE 6. Multi-class classification performance (macro *G − mean* and macro-average *AUC*) of ensemble algorithms.

Model	10-Fold Cross-validation		Test set	
	GM	*AUC*	*GM*	*AUC*
AdaBoost	0.603	0.823	0.608	0.805
GBDT	0.662	0.947	0.676	0.959
RUSBoost	0.665	0.752	0.692	0.793
SMOTE Boost	0.714	0.809	0.732	0.809
Random Forest	0.764	0.939	0.774	0.945
XGBoost	0.777	0.959	0.785	0.967
LightGBM	0.780	0.956	0.788	0.968
Easy Ensemble	0.794	0.839	0.810	0.854
SMOTE Bagging	0.803	0.949	0.852	0.955
DUBE	0.828	0.951	0.869	0.967
Balanced RF	0.873	0.954	0.895	0.968
SPE	0.881	0.949	0.915	0.973
SPE(XGBoost)	**0.888**	**0.963**	**0.923**	**0.976**

However, RUSBoost performs poorly due to the loss of potentially high-quality samples.

3) Comparing the experimental results in Table 5. and Table 6. SPE using XGBoost as base learner shows the highest macro $G-mean$ and macro-average AUC. This shows the good generalization ability of SPE, as well as the potential to deal with imbalanced multi-class classification problems.

The number of base classifiers is crucial to the overall performance of the ensemble learning model. Based on the above experiments, the number of bins is set to 5 (according to [16]) to compare the effects of different base classifiers on the final model. The comparison of the number of base classifiers is shown in Figure 2, taking the average of 100 independent experiments for each iteration. With the increase in the number of base classifiers increases, the macro $G-mean$ of MLP, AdaBoost, Decision Tree, Random Forest, GBDT, LightGBM, and XGBoost generally increases, and the macro $G-mean$ of SVM, KNN, and Logistic Regression does not change significantly. When the number of base classifiers exceeds 20, the macro $G-mean$ of XGBoost, LightGBM, GBDT, Decision Tree, and Random Forest remains stable. After the models converges, SPE using XGBoost as the base classifier has the best performance.

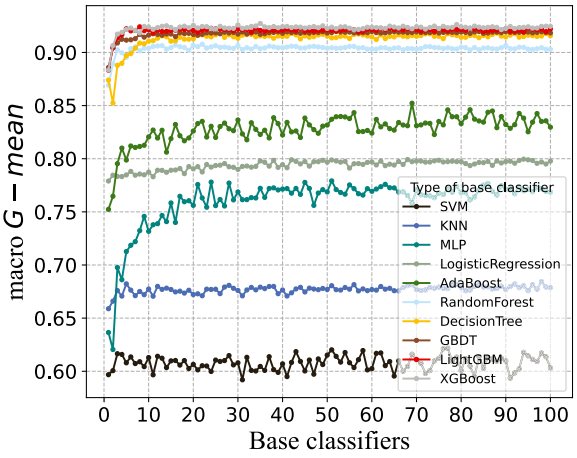

FIGURE 2. Performance on the test set of different base classifiers ranging from 1 to 100.

The testing performance of the SPE algorithm is depicted in Figure 3. The results show that the RSW quality detected by the model has high classification performance, where 57 of the 61 pseudo soldering nuggets were correctly identified, 1086 of 1309 suitable welding nuggets were correctly identified, 82 of 85 expulsion nuggets were correctly identified, and the value of macro $G-mean$ is 0.923 and the value of macro-average AUC is 0.976.

B. FEATURE IMPORTANCE

To analyze the main welding parameters that affect the quality of the weld nugget, SHAP [19] is introduced in this paper. Combining the trained classifier f and an explanation model g, SHAP could calculate importance scores of each feature of every sample, as formulation:

$$g\left(z'\right) = \varphi_0 + \sum_{i=1}^{M} \varphi_i z_i' \tag{7}$$

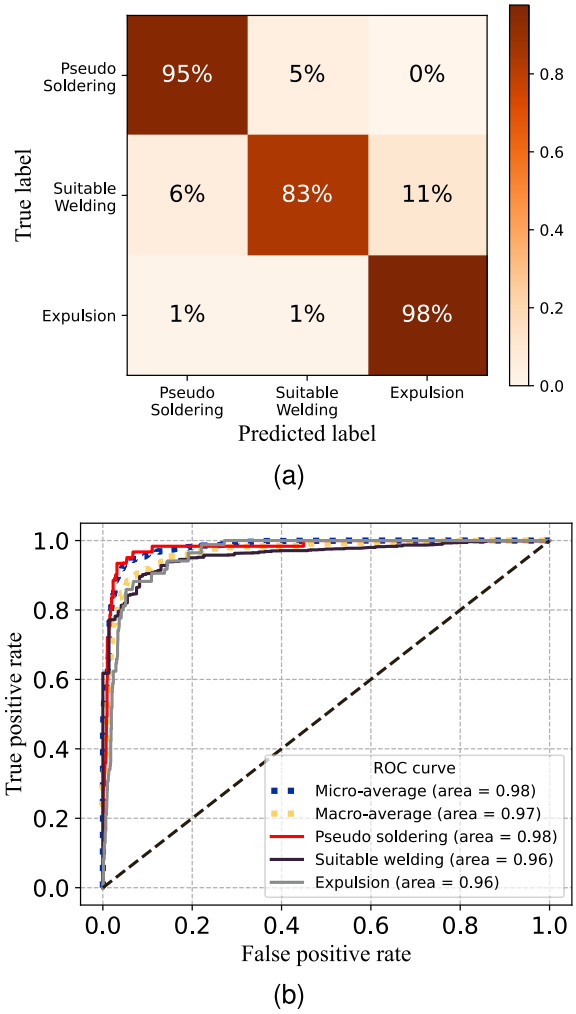

FIGURE 3. Testing performance of quality level identification (a) Confusion matrix. (b) ROC curve.

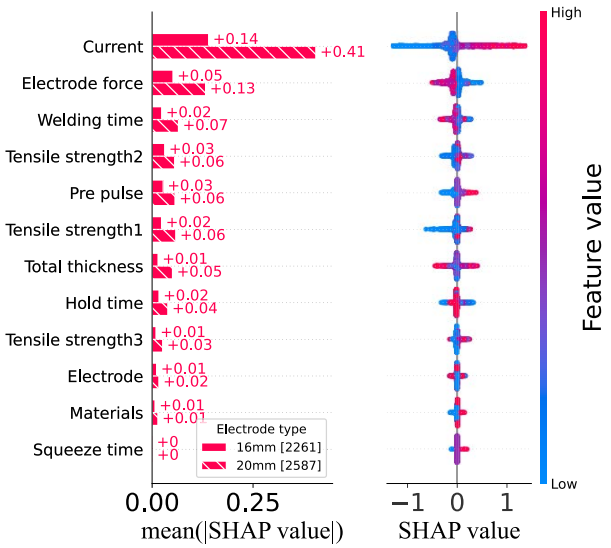

FIGURE 4. Impact on model output.

TABLE 7. Accuracy and error for the models on the dataset 2.

Model	10-Fold Cross-validation	
	R^2	$RMSE$
Lasso	0.083	11.624
MLP	0.207	10.676
SVR	0.277	10.406
AdaBoost	0.910	3.376
NGBoost	0.927	2.970
LightGBM	0.933	2.659
DecisionTree	0.951	1.954
GBDT	0.952	2.542
RandomForest	0.977	1.591
KNN	0.978	1.495
TrAdaBoost.R2(XGBoost)	**0.991**	**1.054**

where M is the number of simplified input features, each feature with a weight φ_i, $z' \epsilon \{0, 1\}^M$, the linear sum of the features is equal to the approximation of f according to local accuracy.

The impacts of each input feature on the nugget quality levels identification are depicted in Figure 4, from a global perspective. The vertical axis of Figure. 4 is the features input to the classification model, sorted from top to bottom according to the degree of influence on the model. The horizontal axis represents the mean value of the absolute SHAP values and the SHAP values. If the SHAP value is a positive number, which means that this feature has a positive impact on the outputs, and conversely, this means that this feature hurts the outputs. Red and blue represent the high and low values of the features, respectively. Each point represents a welding instance.

The results show that the welding current, electrode force, and welding time are the main factors influencing the model. When the current increases, its SHAP value increases, indicating that the larger the current, the greater the positive influence on the model. Therefore, the current is an important parameter to determine the quality of the welding nugget. In contrast, when the electrode pressure takes a higher value, it negatively affects the model. In this case, the welding area increases, so the current density and total resistance are reduced, resulting in heat loss, so the size of the nugget will decrease, and in severe cases pseudo soldering will occur.

C. TENSILE SHEAR STRENGTH PREDICTION

In this subsection, to evaluate the proposed approach, TrAdaBoost.R2 use XGBoost as base learner, we compared the algorithms applied in previous researchs, which include Lasso [4], MLP [4], [10], [53], SVR [4], [53], AdaBoost [4], Decision Tree [4], [12], [53], Random Forest [4], [12], [53], KNN [4], [53], GBDT, LightGBM, and NGBoost [54]. We performed a 10-fold cross-validation on dataset 2, and set default hyperparameters for all models. We repeated the

TABLE 8. Accuracy and error for the models on the dataset 2.

Metric	Model	Training size (%)							
		90	80	70	60	50	40	30	20
R^2	Lasso	0.087	0.100	0.098	0.101	0.098	0.098	0.095	0.092
	SVR	0.282	0.268	0.254	0.248	0.222	0.200	0.175	0.140
	MLP	0.171	0.369	0.386	0.372	0.301	0.334	0.312	0.293
	KNN	0.977	0.974	0.945	0.922	0.831	0.718	0.596	0.438
	LightGBM	0.938	0.938	0.930	0.912	0.881	0.814	0.735	0.567
	DecisionTree	0.944	0.948	0.927	0.924	0.901	0.832	0.748	0.665
	NGBoost	0.929	0.921	0.922	0.918	0.892	0.872	0.824	0.691
	RandomForest	0.970	0.967	0.962	0.946	0.925	0.884	0.834	0.756
	GBDT	0.945	0.944	0.942	0.927	0.920	0.892	0.859	0.789
	AdaBoost	0.913	0.919	0.922	0.920	0.919	0.904	0.884	0.820
	TR	**0.990**	**0.991**	**0.989**	**0.986**	**0.981**	**0.975**	**0.967**	**0.952**
$RMSE$	Lasso	11.880	11.928	11.889	11.879	11.954	11.943	11.970	11.933
	SVR	10.625	10.785	10.826	10.872	11.113	11.250	11.427	11.611
	MLP	10.459	9.872	9.965	10.066	10.552	10.218	10.431	10.583
	KNN	1.528	1.845	2.685	3.271	5.024	6.577	7.928	9.338
	LightGBM	2.725	2.979	3.211	3.600	4.287	5.364	6.451	8.219
	DecisionTree	1.967	2.124	2.612	2.830	3.420	4.569	5.764	6.701
	NGBoost	2.939	3.367	3.505	3.514	3.980	4.401	5.175	6.875
	RandomForest	1.729	2.000	2.255	2.734	3.261	4.077	4.973	6.031
	GBDT	2.494	2.778	2.892	3.204	3.458	3.996	4.604	5.605
	AdaBoost	3.431	3.479	3.425	3.490	3.536	3.823	4.176	5.121
	TR	**1.084**	**1.113**	**1.231**	**1.372**	**1.637**	**1.858**	**2.206**	**2.692**

experiment 100 times respectively, and the average results of R^2 and *RMSE* are reported in Table 7.

To check how the model learns the rules and extends to unknown data sets, we randomly divide the welding dataset 2 into two parts in the ratio of 90:10. 90% of the data samples are randomly selected to train the model, and the remaining 10% of the data is used for model validation. Furthermore, gradually reduce the amount of data in the training set and increase the number of samples in the test set, 80: 20, 70: 30, 60: 40, 50: 50, 40: 60, 30: 70, and 20: 80 are used as the ratio between the training set and the test set to divide the data set. All models use the default parameters, fit the training set 100 times independently, and take the average value of the testing results, which is summarized in Table 8. The visualization of evaluation index comparison on the test set is shown in Figure 5.

As shown in the cross-validation results in Table 7, the R^2 values of the three models of Lasso, MLP, and SVR are less than 0.5, and the *RMSE* values of these three models are more than 11 *KN*. Such precision is unacceptable in the automotive industry. The TrAdaBoost.R2(XGBoost) model performed best with an R^2 value of 0.991 and an *RMSE* value of 1.054 *KN*.

As shown in Table 8, regardless of the data ratio used for training the model, the R^2 values of Lasso, SVR, and MLP are always lower than 0.5, and the prediction error is large, which indicates that these three models are not suitable for predicting the tensile shear strength of RSW. When the proportion of the training set is 60% or more, the KNN, LightGBM, DecisionTree, NGBoost, RandomForest, GBDT, and AdaBoost can maintain high accuracy, When the data used for training is further reduced, the accuracy of these models decreases gradually and the generalization ability weakens.

Regardless of the proportion of data used for training, the TrAdaBoost.R2 (XGBoost) model consistently performs best. In the case of less training data, the TrAdaBoost.R2 (XGBoost) model still maintains high accuracy. When 90 pieces of data (accounting for 20% of the total data set) are used to train the model and 364 pieces of data (accounting for 80% of the total data set) are used for the test, TR performs best among all models, with an R^2 value of 0.952 and an *RMSE* value of 2.692 *KN*. Generally, the above other models perform better on unknown data when the amount of training data is larger. In contrast, the TR model performs well on small samples because enough knowledge is learned from the source domain.

IV. DISCUSSION

This work reports our investigation of the assessment of the quality of resistance spot welding using the data-driven approach.

Since the classification hardness distribution describes the classification difficulty of samples, the sampling weights for the majority class (such as suitable welding, expulsion) can be updated dynamically and potentially useful information

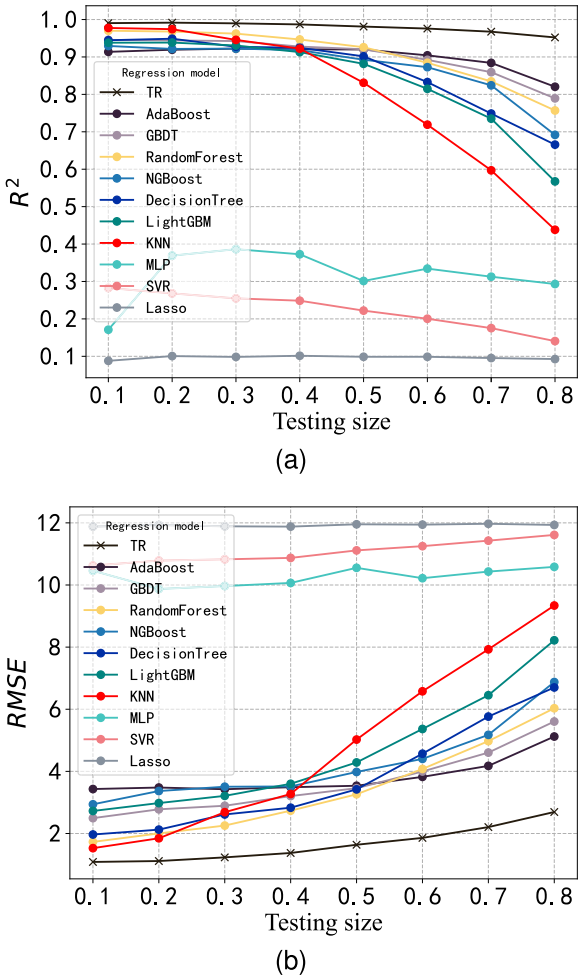

FIGURE 5. Comparison of R^2 and *RMSE* in test dataset for the models (a) R^2 on test set. (b) *RMSE* on test set.

can be maintained. There is inevitably noise in the process of RSW, while the samples generated by over-sampling may be noisy to the dataset. The self-paced factor can reduce the impact of noise samples on the training by reducing the sampling weight of bins with too many samples.

The process parameters required for resistance spot welding using different materials vary widely, and the data used in this study included a large number of dissimilar metal sheets. The inconsistency in the production environment and differences in materials did not influence the performance of the transfer learning model. The proposed tensile shear strength prediction model relies on the availability of historically labeled samples, where a portion of the source domain data is not helpful for the target domain. We will try to handle this problem from the perspective of optimizing the sample weight update strategy.

In the future, we will add more material properties to enrich the feature space. More material properties (such as chemical composition, mechanical properties, and material coating.) may be fused with process parameters as input features of the model to improve the accuracy of quality level identification. Furthermore, the transfer learning method proposed in this paper could be combined with the swarm intelligence method, which in turn further optimizes the welding process parameters, such as reducing the welding current on the premise of ensuring the welding quality.

V. CONCLUSION

Although data-driven machine learning algorithms can be widely used in welding quality prediction, the imbalance of data categories makes it difficult for ML models to distinguish rare welding defects, and limited training data may lead to over-fitting of ML models, which may mislead further process parameter design in the future. To solve these problems, a classification algorithm based on under-sampling technology and ensemble learning is proposed to predict the quality level of RSW. Combining machine learning with the SHAP exPlanations to evaluate the global and local feature contributions, and then using the transfer learning algorithm (TrAdaboost.R2) to learn a model, to predict the tensile shear strength of nugget with a small number of samples. For the application of the proposed methods, the experimental results show that:

1) The SPE identified 95% of pseudo soldering samples, 95% of expulsion samples, and 83% of suitable welding samples on the test set, proving SPE has the best performance among all methods and is suitable for identifying nugget quality levels. This method can assist in quality monitoring during RSW.
2) The result of the SHAP technique indicated that welding current, electrode force, and welding time are the three most important process parameters affecting the welding quality of RSW, which is consistent with previous research.
3) Conventional ML models (such as the random forest) need more training data to keep high accuracy. Due to knowledge extracted from the source domain, the TrAdaBoost.R2 can generalize the knowledge learned from a small number of training samples to unseen samples, and there is no underfitting or overfitting observed in the experiments.
4) The TrAdaBoost.R2 model can assist the design of process parameters without a lot of expensive destructive experiments to verify the weldability.

REFERENCES

[1] Y.-J. Xia, Z.-W. Su, M. Lou, Y.-B. Li, and B. E. Carlson, "Online precision measurement of weld indentation in resistance spot welding using servo gun," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, pp. 4465–4475, Jul. 2020.
[2] B. R. Miao, Y. X. Luo, Q. M. Peng, Y. Z. Qiu, H. Chen, and Z. K. Yang, "Multidisciplinary design optimization of lightweight carbody for fatigue assessment," *Mater. Des.*, vol. 194, Sep. 2020, Art. no. 108910.
[3] X. Cao, Z. Li, X. Zhou, Z. Luo, and J. Duan, "Modeling and optimization of resistance spot welded aluminum to Al-Si coated boron steel using response surface methodology and genetic algorithm," *Measurement*, vol. 171, Feb. 2021, Art. no. 108766.
[4] J. Sim and K.-Y. Kim, "Instance selection-based dissimilar weldment design prediction for resistant spot welding," *Adv. Eng. Informat.*, vol. 52, Apr. 2022, Art. no. 101597.

[5] M. Xiao, B. Yang, S. Wang, Z. Zhang, X. Tang, and L. Kang, "A feature fusion enhanced multiscale CNN with attention mechanism for spot-welding surface appearance recognition," *Comput. Ind.*, vol. 135, Feb. 2022, Art. no. 103583.

[6] W. Dai, D. Li, D. Tang, Q. Jiang, D. Wang, H. Wang, and Y. Peng, "Deep learning assisted vision inspection of resistance spot welds," *J. Manuf. Processes*, vol. 62, pp. 262–274, Feb. 2021.

[7] W. Dai, D. Li, Y. Zheng, D. Wang, D. Tang, H. Wang, and Y. Peng, "Online quality inspection of resistance spot welding for automotive production lines," *J. Manuf. Syst.*, vol. 63, pp. 354–369, Apr. 2022.

[8] L. Zhou, T. Zhang, Z. Zhang, Z. Lei, and S. Zhu, "Monitoring of resistance spot welding expulsion based on machine learning," *Sci. Technol. Weld. Joining*, vol. 27, no. 4, pp. 292–300, May 2022.

[9] A. Dejans, O. Kurtov, and P. Van Rymenant, "Acoustic emission as a tool for prediction of nugget diameter in resistance spot welding," *J. Manuf. Processes*, vol. 62, pp. 7–17, Feb. 2021.

[10] E. Ghafarallahi, G. H. Farrahi, and N. Amiri, "Acoustic simulation of ultrasonic testing and neural network used for diameter prediction of three-sheet spot welded joints," *J. Manuf. Processes*, vol. 64, pp. 1507–1516, Apr. 2021.

[11] S. Kim, I. Hwang, D.-Y. Kim, Y.-M. Kim, M. Kang, and J. Yu, "Weld-quality prediction algorithm based on multiple models using process signals in resistance spot welding," *Metals*, vol. 11, no. 9, p. 1459, Sep. 2021.

[12] S. K. Dang and K. Singh, "Predicting tensile-shear strength of nugget using M5P model tree and random forest: An analysis," *Comput. Ind.*, vol. 124, Jan. 2021, Art. no. 103345.

[13] X. Wan, Y. Wang, D. Zhao, and Y. Huang, "A comparison of two types of neural network for weld quality prediction in small scale resistance spot welding," *Mech. Syst. Signal Process.*, vol. 93, pp. 634–644, Sep. 2017.

[14] H. Zhang, Y. Hou, J. Zhao, L. Wang, T. Xi, and Y. Li, "Automatic welding quality classification for the spot welding based on the Hopfield associative memory neural network and Chernoff face description of the electrode displacement signal features," *Mech. Syst. Signal Process.*, vol. 85, pp. 1035–1043, Feb. 2017.

[15] X. Xie, H. Liu, D. Chen, M. Shu, and Y. Wang, "Multilabel 12-lead ECG classification based on leadwise grouping multibranch network," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[16] Z. Liu, W. Cao, Z. Gao, J. Bian, H. Chen, Y. Chang, and T.-Y. Liu, "Self-paced ensemble for highly imbalanced massive data classification," in *Proc. IEEE 36th Int. Conf. Data Eng. (ICDE)*, Apr. 2020, pp. 841–852.

[17] D. Pan, H. Li, and S. Wang, "Transfer learning-based hybrid remaining useful life prediction for lithium-ion batteries under different stresses," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–10, 2022.

[18] D. Pardoe and P. Stone, "Boosting for regression transfer," in *Proc. ICML*, 2010, pp. 1–8.

[19] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–10.

[20] R. Parihar and S. Jathar, "Grey relational analysis to optimize welding parameters for dissimilar sheets of material in resistance spot welding," *Technology*, vol. 6, no. 11, pp. 23–32, 2015.

[21] T. M. Vinoth and M. Saravanan, "Parametric study in spot welding for dissimilar weld joints," *Int. J. Innov. Res. Sci., Engineeing Technol.*, vol. 5, no. 6, pp. 10704–10711, 2016.

[22] K. Vignesh, A. E. Perumal, and P. Velmurugan, "Optimization of resistance spot welding process parameters and microstructural examination for dissimilar welding of AISI 316L austenitic stainless steel and 2205 duplex stainless steel," *Int. J. Adv. Manuf. Technol.*, vol. 93, nos. 1–4, pp. 455–465, Oct. 2017.

[23] Ó. Martín, P. D. Tiedra, M. López, M. San-Juan, C. García, F. Martín, and Y. Blanco, "Quality prediction of resistance spot welding joints of 304 austenitic stainless steel," *Mater. Des.*, vol. 30, no. 1, pp. 68–77, Jan. 2009.

[24] M. M. Rawal, R. Kolhapure, S. S. Sutar, and V. D. Shinde, "Optimization of resistance spot welding of 304 steel using GRA," *Int. J.*, vol. 3, no. 9, pp. 492–499, 2016.

[25] N. Jadhav, R. Patil, and R. Patil, "Parametric optimization of spot welding metal by Taguchi approach," *Int. J. Eng. Develop. Res.*, vol. 3, no. 4, pp. 857–860, 2015.

[26] S. M. Hamidinejad, F. Kolahan, and A. H. Kokabi, "The modeling and process analysis of resistance spot welding on galvanized steel sheets used in car body manufacturing," *Mater. Des.*, vol. 34, pp. 759–767, Feb. 2012.

[27] H. Sun, J. Yang, and L. Wang, "Resistance spot welding quality identification with particle swarm optimization and a kernel extreme learning machine model," *Int. J. Adv. Manuf. Technol.*, vol. 91, nos. 5–8, pp. 1879–1887, Jul. 2017.

[28] M. Raut and V. Achwal, "Optimization of spot welding process parameters for maximum tensile strength," *Int. J. Mech. Eng. Robot. Res.*, vol. 3, no. 4, pp. 507–517, 2014.

[29] A. Butt and I. Saleem, "Optimization of spot welding processes in low carbon hot rolled sheets," in *Proc. 19th World Conf. Non-Destructive Test.*, 2016, pp. 1–7.

[30] M. Pradeep, N. Mahesh, and R. Hussain, "Process parameter optimization in resistance spot welding of dissimilar thickness materials," *Int. J. Mech. Mechtron. Eng.*, vol. 8, no. 1, pp. 80–83, 2014.

[31] S. Shafee, B. B. Naik, K. Sammaiah, and M. Mohinoddin, "RSW process parameters optimization by Taguchi method," *IOSR J. Mech. Civil Eng.*, vol. 11, no. 5, pp. 46–54, 2014.

[32] F. A. Ghazali, Y. H. Manurung, and M. A. Mohamed, "Multi-response optimization using Taguchi method of resistance spot welding parameters," in *Applied Mechanics and Materials*, vol. 660. Switzerland: Trans Tech, 2014, pp. 120–124.

[33] P. S. Salke and K. C. Bhosale, "Optimization of process parameter in resistance spot welding for unequal thickness sheet using the grey relational analysis," *Int. Res. J. Eng. Technol. (IRJET)*, vol. 3, no. 8, pp. 908–912, 2016.

[34] A. G. Thakur and V. M. Nandedkar, "Optimization of the resistance spot welding process of galvanized steel sheet using the Taguchi method," *Arabian J. Sci. Eng.*, vol. 39, no. 2, pp. 1171–1176, Feb. 2014.

[35] A. G. Thakur, T. E. Rao, M. S. Mukhedkar, and V. M. Nandedkar, "Application of Taguchi method for resistance spot welding of galvanized steel," *ARPN J. Eng. Appl. Sci.*, vol. 5, no. 11, pp. 22–26, 2010.

[36] H. E. Emre and R. Kaçar, "Development of weld lobe for resistance spot-welded TRIP800 steel and evaluation of fracture mode of its weldment," *Int. J. Adv. Manuf. Technol.*, vol. 83, nos. 9–12, pp. 1737–1747, Apr. 2016.

[37] H.-L. Lin, T. Chou, and C.-P. Chou, "Optimization of resistance spot welding process using Taguchi method and a neural network," *Experim. Techn.*, vol. 31, no. 5, pp. 30–36, Sep. 2007.

[38] Y. Zhang, H. C. Chen, Y. Du, M. Chen, J. Liang, J. Li, X. Fan, and X. Yao, "Power transformer fault diagnosis considering data imbalance and data set fusion," *High Voltage*, vol. 6, no. 3, pp. 543–554, Jun. 2021.

[39] M. Kumar, B. Packer, and D. Koller, "Self-paced learning for latent variable models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 23, 2010, pp. 1–9.

[40] D. T. Do and N. Q. K. Le, "Using extreme gradient boosting to identify origin of replication in Saccharomyces cerevisiae via hybrid features," *Genomics*, vol. 112, no. 3, pp. 2445–2451, May 2020.

[41] O. Martin, V. Ahedo, J. I. Santos, and J. M. Galán, "Comparative study of classification algorithms for quality assessment of resistance spot welding joints from pre- and post-welding inputs," *IEEE Access*, vol. 10, pp. 6518–6527, 2022.

[42] J. Kuang, G. Xu, T. Tao, and Q. Wu, "Class-imbalance adversarial transfer learning network for cross-domain fault diagnosis with imbalanced data," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[43] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 28, pp. 321–357, Jun. 2006.

[44] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jun. 2008, pp. 1322–1328.

[45] J. Laurikkala, "Improving identification of difficult small classes by balancing class distribution," in *Proc. Conf. Artif. Intell. Med. Eur.* Berlin, Germany: Springer, 2001, pp. 63–66.

[46] S. Peng, "A GBDT based quality prediction method for the resistance spot welding," in *Proc. Int. Conf. Secur., Pattern Anal., Cybern. (SPAC)*, Jun. 2021, pp. 451–455.

[47] J. Zeng, B. Cao, and R. Tian, "Quality monitoring for micro resistance spot welding with class-imbalanced data based on anomaly detection," *Appl. Sci.*, vol. 10, no. 12, p. 4204, Jun. 2020.

[48] N. V. Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer, "SMOTEBoost: Improving prediction of the minority class in boosting," in *Proc. Eur. Conf. Princ. Data Mining Knowl. Discovery*. Berlin, Germany: Springer, 2003, pp. 107–119.

[49] X.-Y. Liu, J. Wu, and Z.-H. Zhou, "Exploratory undersampling for class-imbalance learning," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 2, pp. 539–550, Apr. 2009.

[50] C. Chen, A. Liaw, and L. Breiman, *Using Random Forest to Learn Imbalanced Data*, vol. 110. Berkeley, CA, USA: Univ. California, 2004, p. 24.
[51] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "RUSBoost: A hybrid approach to alleviating class imbalance," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 40, no. 1, pp. 185–197, Jan. 2010.
[52] Z. Liu, P. Wei, Z. Wei, B. Yu, J. Jiang, W. Cao, J. Bian, and Y. Chang, "Towards inter-class and intra-class imbalance in class-imbalanced learning," 2021, *arXiv:2111.12791*.
[53] F. Ahmed, N.-E. Jannat, D. Schmidt, and K.-Y. Kim, "Data-driven cyber-physical system framework for connected resistance spot welding weldability certification," *Robot. Comput.-Integr. Manuf.*, vol. 67, Feb. 2021, Art. no. 102036.
[54] T. Duan, A. Avati, D. Y. Ding, K. K. Thai, S. Basu, A. Y. Ng, and A. Schuler, "NGBoost: Natural gradient boosting for probabilistic prediction," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 2690–2700.

PEIJIN GUO was born in Wuxue, China, in November 1998. He is currently pursuing the M.S. degree with the Huazhong University of Science and Technology. His research interests include machine learning and software engineering.

QINMIAO ZHU received the M.S. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2014, where he is currently pursuing the Ph.D. degree with the School of Mechanical Science and Engineering. His current research interests include data-based modeling, machine learning, and machine vision.

JINGRAN KANG received the Ph.D. degree from the School of Optical and Electronic Information, Huazhong University of Science and Technology, Wuhan, China, in 2010. He is currently working as a Senior AI Expert at Geely Automobile Research Institute (Ningbo) Company Ltd. His current research interest includes the research and application of AI in manufacturing.

YUHUI WANG is currently a Professor with the State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology. His research interests include intelligent manufacturing, intelligent sensing and RFID, intelligent systems, edge computing, and the Internet of Things. He was awarded the Second Prize of National Technological Invention, in 2013.

WENQIANG HU received the Ph.D. degree from the School of Mechanical Science and Technology, Wuhan, China, in 2005. Since then, she has been a Teacher with the Huazhong University of Science and Technology. Her current research interests include data mining, software engineering, and industrial software.