Transfer learning for quality monitoring of resistance spot welding

The International Journal of Advanced Manufacturing Technology
https://doi.org/10.1007/s00170-026-17898-w

ORIGINAL ARTICLE

Transfer learning for quality monitoring of resistance spot welding

Gabriel Antal[1,2] · Giulia Bruno[1,2] · Valentino Razza[1,2] · Manuela De Maddis[1,2]

Received: 11 November 2025 / Accepted: 13 March 2026

Abstract

Resistance Spot Welding (RSW) is a popular technique for joining sheet metals. Due to the involvement of multiple process parameters, ensuring continuous quality assessment is crucial. While machine learning (ML) methods have demonstrated effectiveness in monitoring welding quality, their application is often limited by the high cost and time required to collect sufficient training data. A promising solution is the integration of prior knowledge into the ML process through transfer learning (TL), enabling the development of more generalized models. This study proposes a TL-based methodology for RSW quality monitoring in the case of limited datasets. An experimental campaign was conducted to generate a target domain dataset comprising welding points produced under varying process conditions. A neural network was trained to predict the nugget diameter, which is typically more difficult and time-consuming to measure. Subsequently, TL techniques were employed to transfer knowledge from a model trained to predict the peak load of tensile shear tests to the nugget diameter prediction model. The source domain dataset used for TL included samples obtained under diverse experimental conditions, encompassing different materials, welding parameters, and electrode types. The results demonstrate that TL enhances model generalization and predictive performance across the full range of nugget diameters, including challenging cases with extremely small or large values that typically hinder accurate prediction. Accordingly, the model performance improved by 25%, achieving a mean absolute percentage error of 5.76%. These findings confirm the potential of TL to improve model robustness, particularly when applied to novel experimental setups. The study provides both theoretical and practical contributions, illustrating how laboratory-generated source domain datasets can be effectively leveraged to support quality monitoring in different production setups.

Keywords Resistance spot welding · Machine learning · Model generalization · Process monitoring · Electrode force

1 Introduction

Resistance spot welding (RSW) is the main welding process in several industrial sectors due to its simplicity, low cost, and efficiency [1]. RSW employs simultaneous electrical power and mechanical force to join together two or more metal sheets through a couple of electrodes. One of the main application fields is the automotive industry [2]. For instance, a single modern car contains up to 7000 spot welds [3].

The RSW process is complex in itself, as it involves multiple parameters, such as welding current, time, and electrode force. Moreover, even having optimized process parameters, other factors can still influence welding quality. Uncontrollable components such as electrode wear, material conditions, the welding machine, and human involvement, particularly in semi-automated operations, can cause deviations from the actual process. Therefore, continuous welding quality assessment is essential [4]. It can be performed

✉ Gabriel Antal
gabriel.antal@polito.it

Giulia Bruno
giulia.bruno@polito.it

Valentino Razza
valentino.razza@polito.it

Manuela De Maddis
manuela.demaddis@polito.it

1 Department of Management and Production Engineering, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino 10129, Italy

2 Advanced Joining Technologies Center (J-Tech@PoliTO), Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino 10129, Italy

Published online: 23 March 2026

through destructive testing (DT) or non-destructive testing (NDT).

Destructive testing methods, including chisel tests, tensile shear testing (TST), torsion testing (TT), and cutting samples for microscopic analysis, provide the most accurate and complete assessment of weld quality. Chisel tests, where a chisel is used to separate the welded sheets, allow direct measurement of weld diameter. TST and TT provide mechanical properties (e.g., the peak load and the maximum torque) that directly relate to joint performance. Nevertheless, selecting the suitable DT method is not a trivial task as it influences fracture mode and the properties of the fracture surfaces [5]. Macrography analysis reveals the internal weld structure, including weld shape, heat-affected zones, and the presence of defects such as porosities, cracks, expulsion, or incomplete fusion. However, DT methods are inherently destructive, which makes 100% inspection of the joints unfeasible and limits their application to sample-based inspections. This limitation, together with the need for time, skilled operators, and specialized equipment, results in additional time and costs.

Non-destructive testing methods offer the advantage of checking every weld without destroying the part, enabling potentially 100% inspection. Ultrasonic tests detect internal problems and measure weld diameter by analyzing reflected sound waves [6]. Thermography methods can identify process problems and predict weld quality based on heating response characteristics [7]. X-ray methods (e.g., computed tomography) provide detailed images of internal weld structure [8]. Magnetic characterization offers the possibility of measuring the weld diameter based on joint magnetic response [9]. Despite these benefits, NDT has several practical limitations in production environments. Many NDT methods require special equipment that is often expensive to purchase and maintain, expert operators, and controlled environmental conditions. Moreover, they can't always be used because of limited access to the weld joints.

Another way of controlling welding quality has been offered by Industry 4.0. Key technologies, such as Machine Learning (ML) and Internet of Things (IoT), enable online monitoring of process signals, providing a straightforward way to assess joint quality. For example, Podržaj et al. have employed a linear vector quantization neural network system to detect the presence of expulsion, i.e., the ejection of liquid metal, one of the most informative indicators of process instability and possible welding defects [10]. More recently, Kershaw et al. have proposed a model for expulsion detection and RSW monitoring, integrating features gathered from process signals and the melting phase of the coating layer [11]. Xia et al. have provided a physics-informed neural network (PINN) framework that integrates a physics-based process and data-driven models to enhance the out-of-distribution generalization of the expulsion prediction model [12]. Taking into account other quality characteristics, such as nugget diameter and thickness, heat-affected zone (HAZ) diameter, and indentation, Russel et al. have compared different models to test their prediction capacity [4]. Another example is proposed in [13], where Bogaerts et al. have developed a method for the nugget diameter prediction based on a combination of unsupervised deep learning and Gaussian process regression.

Machine learning methods have demonstrated their success in RSW process monitoring. However, they have some limitations. Firstly, obtaining high-quality data is both costly and time-consuming. Moreover, the majority of the methodologies applied in the literature are fixed for a specific material and process environment, which is typically laboratory-controlled [14]. As soon as some of the process variables change, the existing models are generally not applicable. A possible way to overcome these limitations is to integrate prior knowledge into the ML process to reach more generalized models. Different types of knowledge representations (e.g., algebraic equations, differential equations, simulation results, logic rules, and human feedback) can be incorporated into the ML pipeline, leading to the concept of informed machine learning [15]. For instance, Physics-Informed Neural Networks (PINNs) integrate physics-based scientific knowledge into neural networks. A further solution to enhance model generalization comes from transfer learning (TL). TL reuses a previous model to assist the new one, which eventually deals with different process conditions (e.g., materials, input parameters, etc.) [16]. In other words, it can be stated that TL enables the integration of knowledge from one or more source domain ML models to a target domain model [17]. Although TL has the potential to solve many issues for RSW, there are just a few examples in the literature. Guo et al. have proposed a TL method based on *TrAdaBoost.R2* and *XGBoost* that employs process and material parameters to predict the tensile shear strength [18]. Noh et al. have used a TL-based convolutional neural network (CNN) fed by current, voltage, and acceleration signals to classify normal from abnormal spot welds [19]. Using images instead of process parameters or signals is popular in quality monitoring. Xiao et al. have proposed a CNN method to recognize spot welding appearances [20]. Similarly, a TL-based CNN architecture that predicts the nugget diameter from thermal images has been proposed by Santoro et al. [21]. Another solution to deal with data scarcity comes from Yue et al. The authors first trained a neural network with low-fidelity data obtained through finite element simulations. Then, they froze some layers to transfer them to the new model trained with experimental data [22].

This work aims to develop a straightforward TL-based methodology for RSW monitoring in cases where a limited dataset is available, reflecting the practical challenge in the industrial world where obtaining labeled data is often costly and time-consuming. First, a neural network has been trained to predict the nugget diameter, which is generally more difficult and time-consuming to measure compared to the TST peak load. Second, TL techniques have been used to transfer the knowledge of a TST predicting model to the nugget one. Lastly, the models have been compared. Most importantly, results show that the TL enhances model generalization. Even in the case of folds containing abnormal nugget diameter dimensions (e.g, the small ones), the error drastically decreases, compared to the non-TL model.

The remainder of the paper is organized as follows. Section 2 describes the applied methodology in this paper. Section 3 discusses the results, while Section 4 deals with conclusions and proposals for future implementation.

2 Methodology

The adopted terminology is taken from Giannetti and Essien [23]. A domain $\mathcal{D} = \{\mathcal{X}, \ell\}$ consists of a feature space $\mathcal{X}$ and ℓ observations. We consider two types of datasets, the source domain and target domain datasets, respectively $D_s = \{(x_{s1}, y_{s1}), ..., (x_{s\ell}, y_{s\ell})\}$ and $D_t = \{(x_{t1}, y_{t1}), ..., (x_{t\ell}, y_{t\ell})\}$. A task T is defined as $T = \{y, \varphi(x)\}$, where $x_i \in X$ are observations in the input space, $y_i \in Y$ is the response, and $\varphi(\cdot)$ is a predictive function. We consider T_s and T_t as the source and the target tasks, respectively. Transfer learning seeks to enhance the learning task T_t in the target domain by using knowledge in D_s and T_s, where $D_s \neq D_t$ or $T_s \neq T_t$, given a source domain $\mathcal{D}_s$, a learning task T_s, a target domain $\mathcal{D}_t$ and a learning task T_t.

The proposed methodology is shown in Fig. 1. The paper structure is based on the same figure. The definition of the target domain model is done in four phases. The problem definition leads to the identification of the target variable (one or more). The data collection describes the experimental campaign and the process of gathering the raw data. The feature engineering and selection phase is used to analyse the features and rank them to see how the number of involved features affects the results. At the end of these three phases, the target domain dataset (D_t) is defined. Finally, the D_t model building/training/validation phase describes the prediction model for the target domain dataset. After that, the definition of the Transfer learning model involves two additional phases. The first defines the data set of the source domain (D_s), i.e., the additional data and models used to improve knowledge. The TL models building/training/validation phase describes how the target dataset is integrated into the source models to transfer the knowledge. The final phase deals with the comparison between the performance obtained by the transfer learning models and the D_t one.

2.1 Problem definition

The first step involves defining the problem. Our goal is to improve quality evaluation in RSW by employing process signals. Examples of the main RSW quality criteria are the nugget diameter, the TST peak load, the weld diameter, and the indentation. As one of the most desired and costly-to-measure parameters, our target is the nugget diameter. The ISO standard 17677-1-2021 [24] defines the weld nugget diameter d_n as the “diameter of nugget measured at the faying surface by metallurgical examination”. The influence of nugget diameter on the fracture behavior of spot-welded joints is well established in the literature. In general, a reduced nugget diameter leads to interfacial failure, whereas an increased nugget size usually results in plug-type fractures (i.e., failure mode resulting in a plug on one sheet and

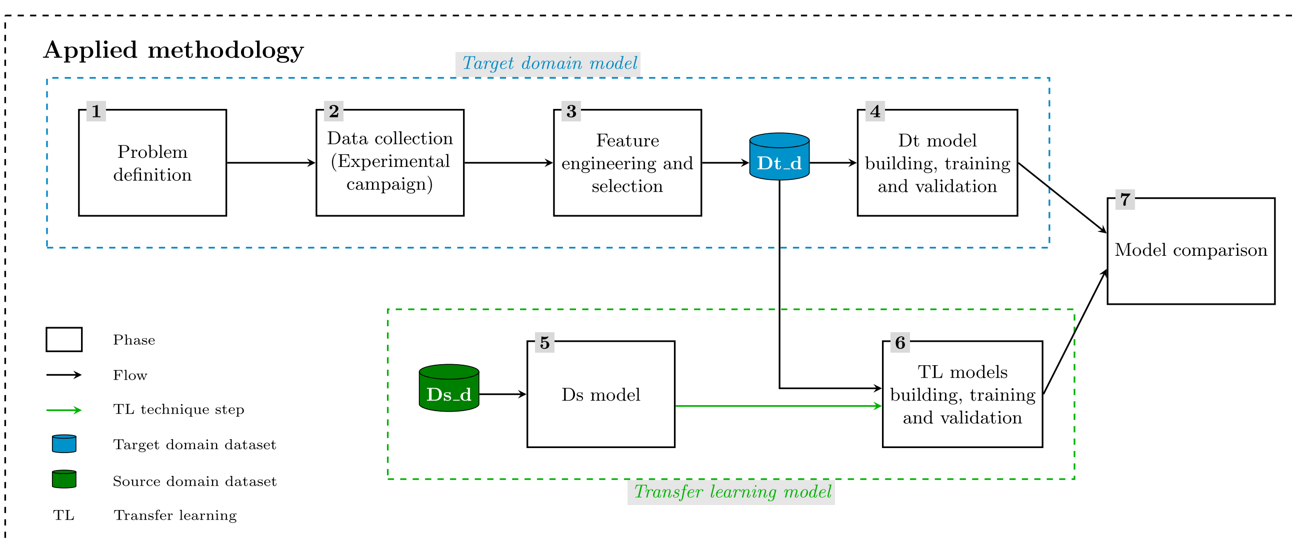


Fig. 1 Methodology adopted

a hole in the other sheet). Of these modes, plug failure is the most desirable, as it is typically associated with higher joint strength [25, 26].

2.2 Data collection - experimental campaign

The experimental campaign has been conducted using a medium-frequency direct current RSW machine, working in "constant current" mode and equipped with a TE700 (Tecna) control unit (Fig. 2). The force signal has been acquired using a piezoelectric surface strain sensor (Kistler Italia, mod. 9232A) through a National Instruments cRIO 9035 at a 40 kHz sampling rate. The material employed is the 1 mm thick GI40/40-U DP590 steel. The specimen dimensions are (45x105) mm, with an overlapping area of 35 mm (Fig. 3a). During the welding, the cooling rate is 4 L min^{-1}. Experiments have been conducted using Cu-Cr-Zr electrodes with a 5 mm contact diameter and a truncated cone shape. The weld time is 250 ms (upslope=25 ms, current time=200 ms, downslope=25 ms). Multiple sets of welding current (variable values between 8 and 14 kA) and welding pressure (variable values between 1 and 2.2 bar) have been chosen to realize 48 spot welds with the aim of having conforming and non-conforming welding points. The experimental setup has been dictated by ISO 14273 and 14373 standards[27, 28]. Metallurgical examination has been utilized to measure the weld nugget diameters (Fig. 3b).

Fig. 2 RSW machine used during the experimental campaigns

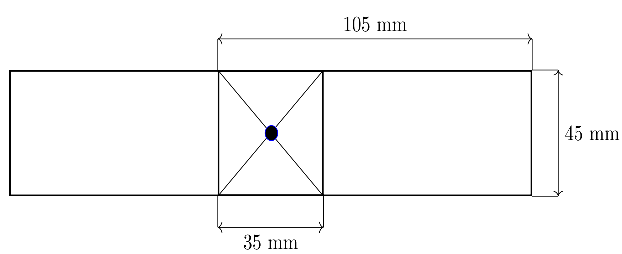
2.3 Feature engineering and selection

Being one of the most informative signals in RSW [10], the electrode force has been employed to predict the nugget diameter. In the next subsections, the force behavior and the feature engineering phase adopted to construct the datasets are explained.

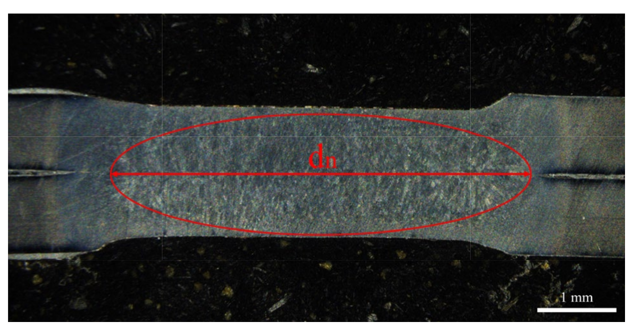
2.3.1 Force curve

Figure 4 displays an example of electrode force and current curves acquired and processed after the experimental phase. The force is the response of the applied input pressure. During the slope-up time (0-25 ms), the welding current starts and reaches the imposed value. The next phase is the current time (25-225 ms), and the current is kept constant and equal to the input current chosen.

The electrode force varies throughout the welding cycle following a physically consistent pattern. Initially, the applied heat causes thermal expansion of the workpieces, thereby increasing the electrode force. As the temperature rises further, the material softens and undergoes plastic deformation, which compensates for the thermal expansion, resulting in a relatively stable force plateau. Toward the end of the process, when the current is reduced (slope-down phase) and finally turned off, the metal cools and contracts,



(a) Specimen dimensions according to ISO 14273.



(b) Weld nugget cross section.

Fig. 3 Specimen dimensions (**a**) and an example of a weld nugget macrography made during the experimental campaign (**b**). The red arrow indicates the weld nugget diameter, d_n

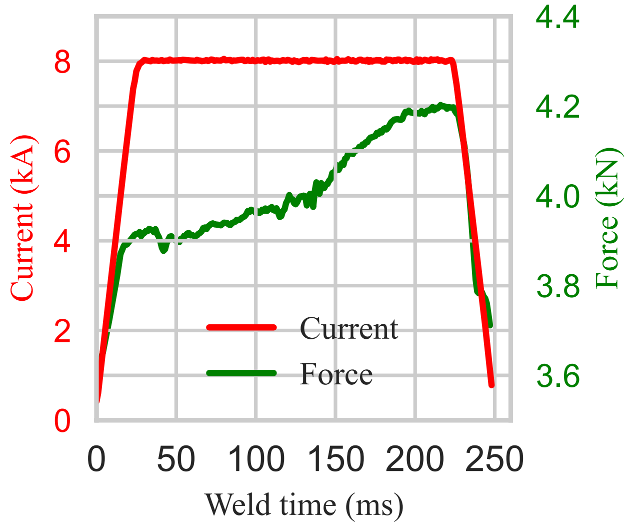


Fig. 4 Example of electrode force and current curves acquired during the experimental campaign

resulting in a decrease in electrode force. The thermal behavior of the electrode force is well-known in the literature [29, 30] and contains core information regarding the welding status used for the feature extraction phase.

2.3.2 Feature engineering

To enhance computational efficiency and improve model performance during training, a feature extraction process has been employed. Leveraging domain knowledge, the most relevant features have been selected to capture the key characteristics of the electrode force curve. Data extracted from the force sensor is contained in two vectors

$$\begin{aligned} T^i &= [t_1^i, \dots, t_j^i, \dots, t_\eta^i] \quad (\mathrm{s}) \\ F^i &= [f_1^i, \dots, f_j^i, \dots, f_\eta^i] \quad (\mathrm{kN}) \end{aligned} \tag{1}$$

where F^i corresponds to the i-th spot weld force vector, f_j^i is the force value acquired at time t_j^i from the beginning of the experiment, and η is the number of values acquired during the welding. T^i is the vector of the sampling times. The features considered for this study are:

- Maximum force value $\overline{f}^i$

$$\overline{f}^i = \max_{j \in \{1,2,\dots,\eta\}} f_j^i \quad (\mathrm{kN}) \tag{2}$$

- Time corresponding to the maximum force t_M^i, where M is

$$M = \arg \max_{j \in \{1,2,\dots,\eta\}} f_j^i \tag{3}$$

- Difference between the maximum force and the first force value, divided by the time interval Δ_1^i

$$\Delta_1^i = \frac{\overline{f}^i - f_1^i}{t_M^i} \quad (\mathrm{kNs}^{-1}) \tag{4}$$

- Difference between the last and the maximum force, divided by the time interval Δ_η^i

$$\Delta_\eta^i = \frac{\overline{f}^i - f_\eta^i}{t_\eta^i - t_M^i} \quad (\mathrm{kNs}^{-1}) \tag{5}$$

- Standard deviation of the force values σ_f^i

$$\sigma_f^i = \sqrt{\frac{1}{\eta - 1} \sum_{j=1}^{\eta} (f_j^i - \mu_f^i)^2} \quad (\mathrm{kN}) \tag{6}$$

- Skewness of the force values γ_f^i

$$\gamma_f^i = \frac{\frac{1}{\eta} \sum_{j=1}^{\eta} (f_j^i - \mu_f^i)^3}{\left(\frac{1}{\eta} \sum_{j=1}^{\eta} \left(f_j^i - \mu_f^i \right)^2 \right)^{3/2}} \tag{7}$$

- Coefficient of variation of the force values c_f^i

$$c_f^i = \frac{\sigma_f^i}{\mu_f^i} \tag{8}$$

- Difference between the third and the first quartiles of the force curve δ_Q^i

$$\delta_Q^i = Q_3(F^i) - Q_1(F^i) \quad (\mathrm{kN}) \tag{9}$$

- Median absolute deviation of the force values, MAD^i

$$\mathrm{MAD}^i = \mathrm{median}(|f_j^i - \mathrm{median}(F^i)|) \quad (\mathrm{kN}) \tag{10}$$

The features mentioned above have been extracted for each spot weld. The final dataset also contains the nominal welding parameters, i.e., input current (I^i) and pressure (P^i), the material thickness (z_{mat}^i), and the welding time information regarding upslope, current time, and downslope, being (t_{up}^i, t_{c}^i, t_{down}^i), respectively. Finally, it is completed by the target variable (i.e., the weld nugget diameter). In total, fifteen features have been identified.

2.3.3 Feature selection

A feature selection (FS) process has been employed to assess the importance of each feature. Specifically, four

Table 1 Feature scores

Feature	Final Score
P^i	0.65
t^i_M	0.63
σ^i_f	0.60
c^i_f	0.50
I^i	0.44
Δ^i_1	0.42
f_{MAD}	0.33
δ^i_Q	0.31
γ^i_f	0.23
Δ^i_η	0.13
$\overline{f}^i$	0.03
t^i_{up}	0.00
t^i_{c}	0.00
t^i_{down}	0.00
z^i_{mat}	0.00

different FS methods have been used. Mutual information (MI) is a measure that quantifies the amount of information about one variable through observing another variable [31]. F-regression is a statistical method used to rank the relevance of features for predicting the target variable by computing an F-statistic for each feature, measuring the linear relationship with the target [32]. Random Forest (RF) feature importance performs feature selection by identifying the most important features, based on their contributions, for predicting the target variable [33]. Lastly, the Spearman rank correlation returns a measure of the monotonicity of the relationship between the features [34].

Each score has been normalized (min-max normalization) and the average score has been computed to obtain the final feature ranking. The final ranking is shown in Table 1. The scores for the welding time information (slopes and current time) and material thickness are not informative, as they remain constant in our experiment. Conversely, they could be important in experiments with variable input parameters and material thickness.

2.4 D_t model building, training and validation

Following the creation of the D_t dataset (48 samples), a model was built to predict the weld nugget diameter (Fig. 5a). The proposed network has only one hidden (Dense) layer. For this reason, it is referred to as a shallow neural network (SNN). SNNs are networks having one or two hidden layers. This architecture has been chosen because SNNs typically outperform deep neural networks (DNNs) for small datasets [35].

The number of hidden neurons, $n = \{4, 6, 8, 10, 12\}$, and the number of batch size, $b = \{1, 4, 8, 16\}$ have been tuned. The Dense layer has a "ReLU" activation function. The optimizer is the "Adam" one while the loss function is the mean absolute percentage error (MAPE). The dataset has been divided into 6 folds in reverse chronological order (fold 1 contains samples 41-48, fold 2 contains samples 33-40, etc.) to ensure that all samples are included in at least one of the training or test sets. Consequently, a 6-fold cross-validation has been used. Moreover, to ensure comparability and reproducibility, a random seed has been set. The model has been trained for 5000 epochs per each fold

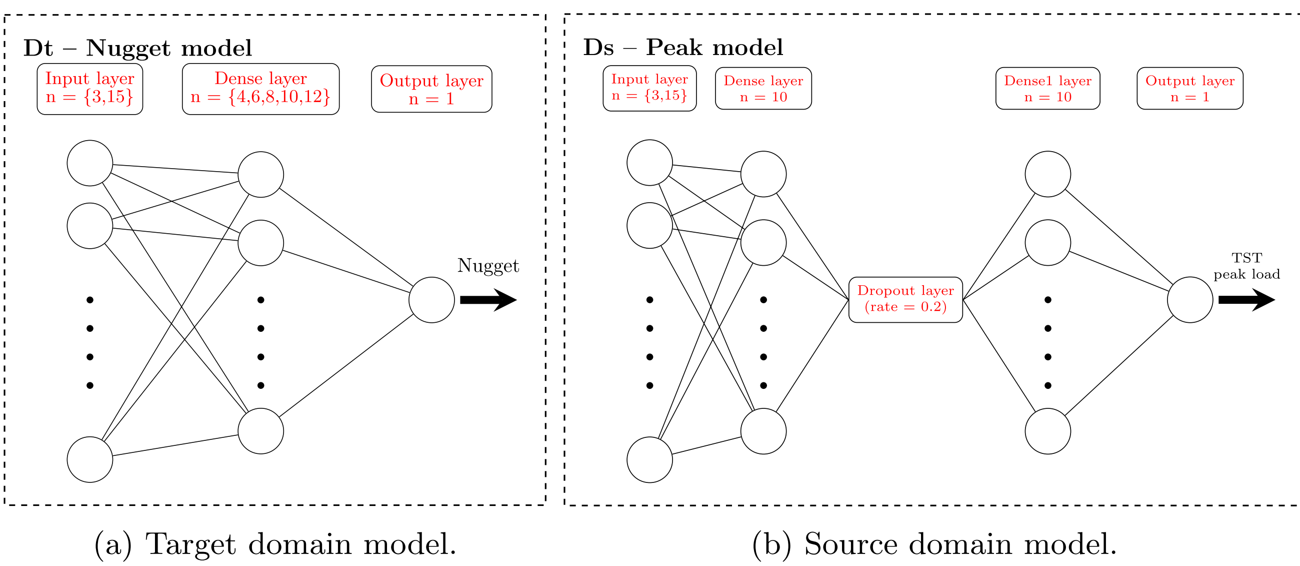


Fig. 5 Target domain (**a**) and source domain (**b**) models. The first predicts the weld nugget while the second predicts the TST peak load. All features and the selected top 3 have been considered for the analysis

and configuration. Thus, the lowest validation MAPE has been taken for that specific fold and configuration (n and b). The best model has the configuration with the lowest average MAPE computed on the 6 folds. Based on the feature selection, optimal performance was achieved when utilizing the three highest-ranked features (P^i, t^i_M, σ^i_f). Moreover, the D_t model has also been run and tuned employing all features. The best MAPE is $7,71\%$ in the case of selected features. A comprehensive analysis of these results will be presented in the corresponding section.

2.5 Source datasets and related model

Before the next step of Fig. 1, a question arises: Are there additional related data that could assist the D_t model in predicting the nugget diameter? If so, this data will form the source dataset D_s. In our case, it comprises 108 samples and is composed of the 3 datasets used in the authors' past studies. The welding machine employed, the welding mode (constant current), the cooling rate, the material type (DP590), and the electrode geometry were the same.

The purpose of campaigns C1 and C2 was to investigate the electrode wear effect. The welding parameters were fixed for each campaign (C1 - upslope=100 ms, current time=300 ms, downslope=100 ms, current=8 kA, force 3 kN, and C2 - current time= 380 ms, current=8 kA, force=2.9 kN). The electrode diameters were 6 mm for C1 and 5 mm for C2, while the material thicknesses were 1 and 0.8 mm. Differently, the purpose of campaign C3 was to investigate the expulsion presence. In this case, the weld time was 200 ms (upslope=25 ms, current time=150 ms, downslope=25 ms) while the input parameters were variable (welding current between 6.5 and 8 kA and welding force between 1.6 and 2.3 kN). The electrode face diameter was 4.5 mm while the sheets thicknesses where 0.8 mm. Table 2 contains brief information about the main differences of the 3 campaigns.

The force curve behaviour discussed in Section 2.3.1 is valid even for D_s. Consequently, the same features shown for D_t have been extracted for D_s. Furthermore, the learning tasks are different ($T_s \neq T_t$). For D_t the target variable is the nugget diameter, while for D_s, it is the TST peak load. The reason for using this D_s dataset is that it could be beneficial for the original learning task (i.e., predicting the nugget diameter) since the weld strength and the nugget diameter are usually highly correlated [26]. For example, Zhao et al. [39] have reported in their study a Pearson correlation of 0.9 between the nugget diameter and the peak load.

The D_s model structure predicting the peak load is shown in Fig. 5b. After the tuning phase in the case of using the top 3 features obtained from the FS process for the target dataset D_t (P^i, t^i_M, σ^i_f), the best results have been obtained with a batch size of $b = 1$, where the losses (MAPE) for the training and validation sets are, respectively 8.86% and 10.52%. As for the D_t model, the D_s model with all features has been constructed. In this case, the training and validation losses are, respectively 8.08% and 7.83%.

2.6 TL models building, training and validation

The next step in the proposed methodology is to construct the transfer learning (TL) models. Since the peak load and the weld nugget diameter are usually highly correlated, we expect that using pre-trained layers from the D_s model (either as initialization or with frozen weights) would be advantageous for learning our target task. The objective is to transfer the knowledge of the D_s model to new models trained on the D_t dataset (48 samples), predict the nugget diameter, and subsequently compare their performance with that of the baseline D_t model presented in Fig. 5a.

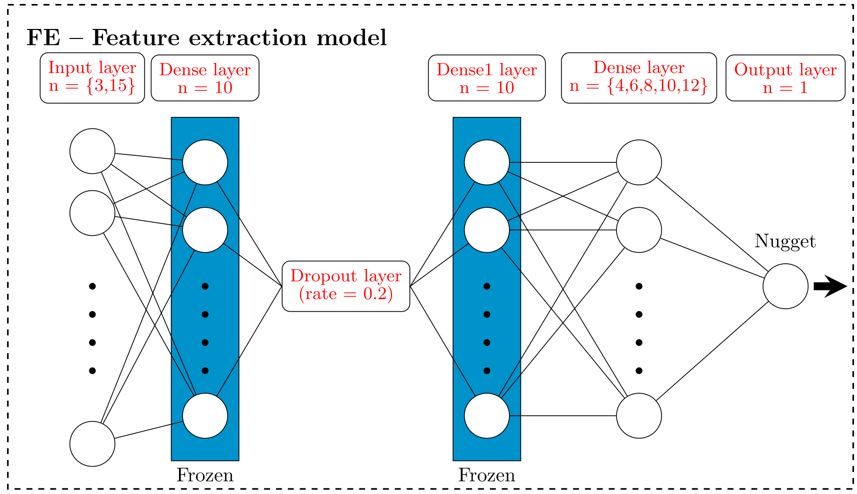
To do so, we take the D_s model, remove the output layer, and add a new dense layer. We consider three cases: feature extraction (FE), fine-tuning (FT), and full fine-tuning (FF-T). The first one consists of freezing the D_s model layers (Fig. 6a). It is called so since it acts like a feature extractor for the new layer, which will update its weights accordingly. The second one has only one frozen layer. Consequently, the other layer gained from the D_s model is free to update its weights (i.e., fine-tuning) with the new training (although initially, it starts from the weights obtained during the D_s model training, Fig. 6b). Lastly, in the third case, all layers are unfrozen (Fig. 6c) and fine-tuned during the training process. Therefore, in terms of hidden layers, the TL models can be considered as the union of the D_t and D_s models.

To compare the TL models with D_t and evaluate the D_s model contribution, the same parameter tuning discussed in Section 2.4 has been applied. Specifically, the number of hidden neurons, $n = \{4, 6, 8, 10, 12\}$, and the number of batch size, $b = \{1, 4, 8, 16\}$ have been tuned for the last hidden layer. Similarly, the same random seed has been set, a 6-fold

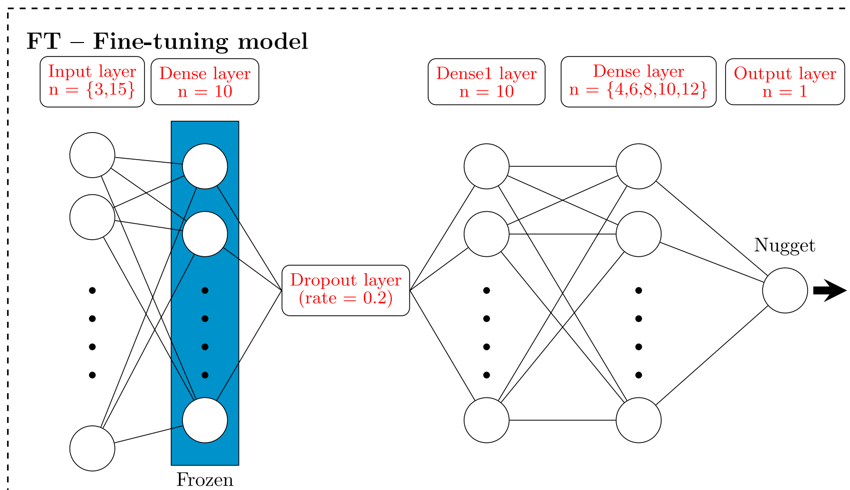
Table 2 Differences of source domain dataset campaigns. For each sample, the TST has been conducted. Material thickness (t), electrode diameter (d_e), input parameters, and the study purpose differ. Detailed information is available in the reported papers

Campaign	t (mm)	d_e (mm)	Parameters	Samples	Purpose	Papers
C1	1	6	constant	30	electrode wear effect	[36, 37]
C2	0.8	5	constant	39	electrode wear effect	[37]
C3	0.8	4.5	variable	39	expulsion presence	[38]

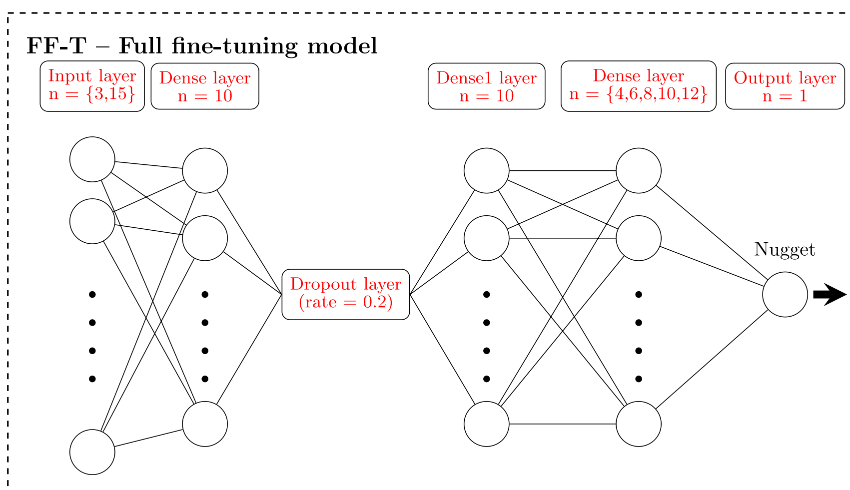
Fig. 6 FE (**a**), FT (**b**), and FF-T (**c**) models. The initial part corresponds to the D_s model. All features and the top 3 have been used. Moreover, as for the D_t model, a different number of neurons has been employed for the last layer



(a) Feature extraction model.



(b) Fine-tuning model.



(c) Full fine-tuning model.

Table 3 Models best average performance comparison: All features vs top 3 selected features (P^i, t^i_M, σ^i_f). The reported MAPE refers to the average computed across the 6 folds

	All Features		Selected Features	
Model	Neurons - Batch	MAPE (%)	Neurons - Batch	MAPE (%)
D_t	4-8	10.28	10-4	7.71
FE	8-4	**6.38**	12-1	**5.76**
FT	12-8	6.90	4-1	7.33
FF-T	4-1	8.44	4-1	6.79

cross-validation has been used, and the average MAPE has been considered to assess the specific performance.

The final phase shown in Fig. 1 (Model comparison) will be discussed in the next section.

3 Results and discussion

As previously explained, we recall that the D_t dataset (48 samples) has been partitioned into 6 folds to ensure comprehensive coverage, with all samples included in either the training or test sets throughout the validation process. Consequently, a 6-fold cross-validation has been employed to evaluate model performance. For each fold and configuration, the model has been trained for 5000 epochs, with the MAPE being monitored throughout the training process. The minimum validation MAPE achieved during training has been considered for each specific fold and parameter configuration (number of neurons and batch size). The best model configuration has been identified by selecting the set of parameters that produces the lowest average MAPE across the 6 folds.

Table 3 shows the results comparing two cases: the first considers models performances fed with all 15 features extracted. In contrast, the second case only considers the top 3 features, which have led to the lowest average MAPE for the *Dt* model. The feature selection importance is confirmed in this case, as the *Dt* model improves from 10.28% to 7.71%. In general, using only (P^i, t^i_M, σ^i_f) improved the models performances. The only exception is the *FT* model, which shows a small worsening (from 6.38% to 7.33%). In both cases, the FE model has outperformed, with the best MAPE = 5.76%, obtained with the top 3 selected features.

Going deeper into the discussion, we focus on the characteristics of the *Dt* dataset (see Section 2.2). Different input parameters have been used to perform 48 samples. Table 4 shows the nugget diameter distribution for each fold. The dataset mean nugget is 5.24 mm with a standard deviation of 0.94 mm while the minimum and the maximum are, respectively, 1.8 mm and 6.7 mm. There are 32 samples with a nugget diameter between 5.0 mm and 5.9 mm, 7 nuggets greater than or equal to 6.0 mm and 9 nuggets below or equal to 5.0 mm. Therefore, the dataset is highly imbalanced, but this is completely normal since the majority of the welding points fall inside the weldability lobe (i.e., range of acceptable welding parameters that produce a spot weld of a desired quality). Moreover, based on ISO 14373 [28] (the desired diameter is 5 mm while the minimum acceptable should be at least 3.5 mm), we identify four categories for the folds: normal, normal-large, mixed, and challenging. The normal one contains diameters around 5 mm and has a low standard deviation. The second category contains some large diameters, while the third one has mixed values (small, normal, and large). The last category (Fold 2) is the most challenging one since it contains the smallest diameters and has a huge standard deviation (i.e., 1.5 mm).

Figure 7 shows the loss curves for each model (best configuration in terms of neurons and batch size) in the case of selected features. Each row corresponds to a fold, while each column corresponds to a model. We focus on folds 2, 3, and 6. As the training MAPE (blue curve) lowers with epoch increases, the test MAPE gets higher, meaning that overfitting occurs for *Dt*, *FT* (one transferred layer free), and $FF-T$ (both transferred layers free) models. The overfitting phenomenon is even more pronounced for folds 2 and 6 compared to fold 3. Indeed, these two folds contain small diameters (1.8 mm - 3.6 mm) while fold 3 contains some large diameters. Consequently, the behavior indicates that the models are unable to generalize. On the other hand, the *FE* model (source domain D_s internal layers frozen, see Fig. 6a) shows a tendency towards convergence in the loss curves, particularly for folds 2 and 3. This behavior suggests that the *FE* model has the capability of generalization in the case of unseen data, thanks to the fixed knowledge integration of the D_s model (peak load prediction, see Section 2.5). In other words, in the case of validation folds containing the

Table 4 Validation fold composition and nugget diameter statistics. Bold values indicate abnormal nugget diameters

Fold	Samples	Min (mm)	Max (mm)	Mean (mm)	Std Dev (mm)	Category
1	41-48	4.90	5.70	5.25	0.28	Normal
2	33-40	**1.80**	5.30	4.36	1.50	**Challenging**
3	25-32	5.18	**6.70**	5.61	0.59	Normal-large
4	17-24	5.30	6.10	5.74	0.29	Normal-large
5	9-16	**3.40**	6.10	5.14	1.03	Mixed
6	1-8	**3.29**	6.24	5.32	0.90	Mixed

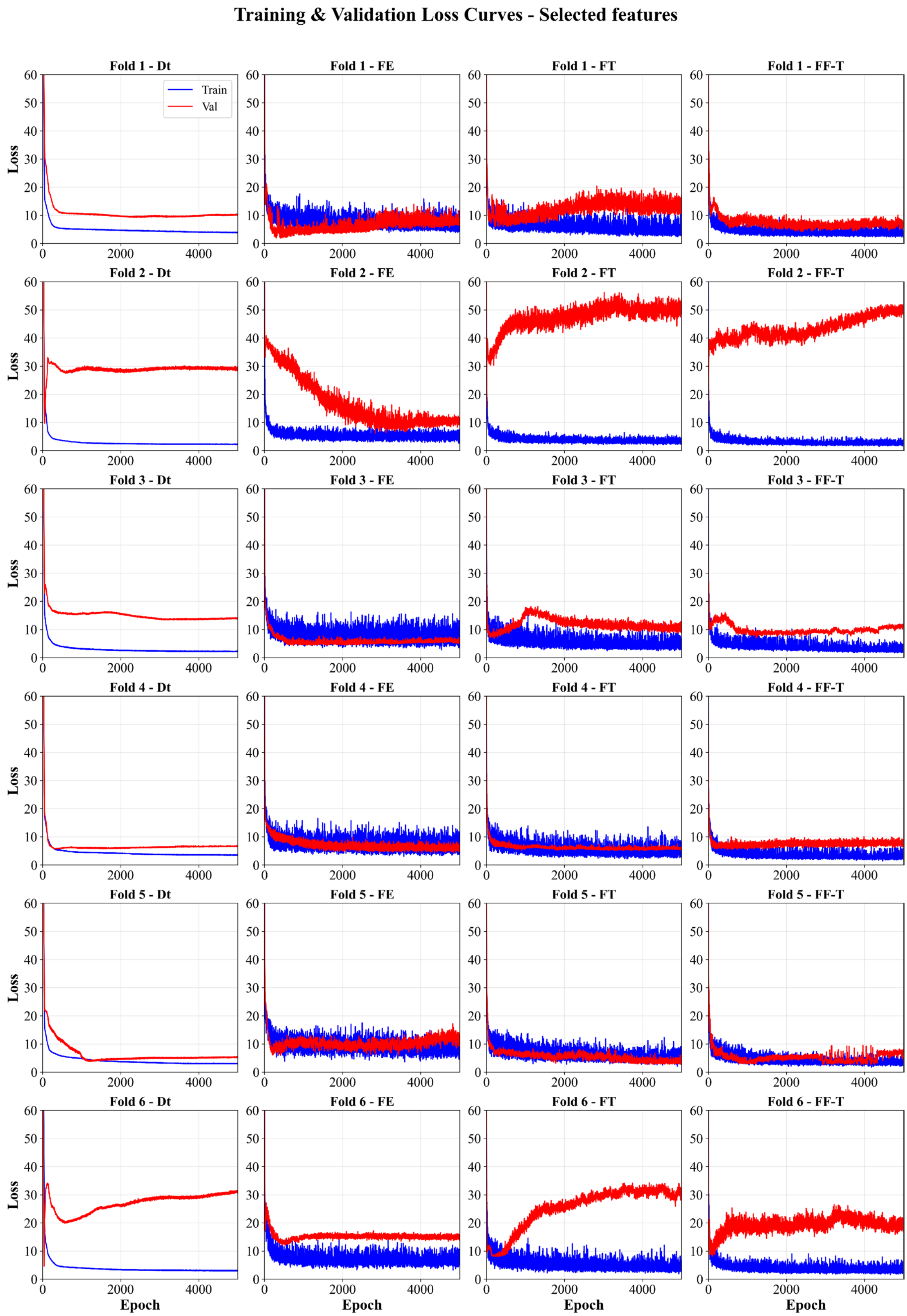


Fig. 7 Models loss (MAPE) curves for the selected features run (P^i, t_M^i, σ_f^i), across 5000 epochs. Each row refers to a fold. The number of neurons of the tuned layer and batch size are respectively D_t(10-4), FE(12-1), FT(4-1), $F-FT$(4-1)

Table 5 Model performance (MAPE %) for each fold using top 3 selected features. The number of neurons and batch size for each model are reported in Table 3. The best results are highlighted in bold

Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Mean	Std Dev
D_t	9.22	9.69	13.40	5.61	3.73	**4.59**	7.71	3.70
FE	**1.87**	**6.14**	**4.30**	**4.34**	5.99	11.92	**5.76**	**3.39**
FT	6.03	15.44	6.81	5.03	2.67	8.00	7.33	4.36
FF-T	4.19	13.14	6.97	5.46	**2.66**	8.11	6.76	3.68

Table 6 Training dynamics: Best epoch, training MAPE (%), and validation MAPE (%) for D_t and *FE* models (neurons-batch). Bold highlights indicate suspicious early convergence before adequate training

Fold	Nugget Range	D_t (10-4)			FE (12-1)		
		Epoch	Train	Val	Epoch	Train	Val
1	4.90–5.70	2408	4.46	9.22	406	11.55	1.87
2	1.80–5.30	**54**	**18.96**	**9.69**	2839	5.51	6.14
3	5.18–6.70	3366	2.23	13.40	742	7.35	4.30
4	5.30–6.10	309	5.59	5.61	4349	5.97	4.34
5	3.40–6.10	1194	4.41	3.73	414	11.08	5.98
6	3.29–6.24	**39**	**39.32**	**4.59**	451	8.77	11.92

extreme nuggets, the D_t model can't learn them anymore, while the transfer model *FE* benefits from the knowledge (i.e., the frozen layers neurons weights) contained in the D_S model that performs a learning task, which has a target variable (i.e., the peak load), generally highly correlated with the nugget diameter. Moreover, it is interesting to note that as soon as one layer or both layers are not frozen, the behavior on the mentioned folds is similar or even worse than that of the D_t model.

Table 5 shows the folds validation MAPE for each model (best configuration in terms of neurons and batch size). By comparing the D_t model with the best TL model (FE), it can be observed that D_t is better for folds 5 and 6. Specifically, for fold 6, the difference is huge, which apparently partially contradicts the TL benefits.

For the reason above, the training dynamics have been deeply analyzed for D_t and FE models. Table 6 reports important differences. Notably, for fold 6, D_t achieved its lowest validation MAPE (4.59%) at epoch 39, when the training MAPE was still 39.32%, indicating the model had barely begun learning the training data. In contrast, FE converged at epoch 451 with a more reasonable train-validation relationship (8.77% vs 11.92%). Similarly, fold 2 shows D_t stopping at epoch 54 with high training error (18.96%), while FE converges at epoch 2839 with well-balanced train-validation performance (5.51% vs 6.14%). These patterns suggest that D_t reported performance (best MAPE) may be optimistically reached due to random points, whereas FE demonstrates more stable convergence behavior. The frozen transferred layers appear to provide implicit regularization that guides the model toward more robust results, reducing sensitivity to the random initialization that could affect shallow networks trained on small datasets.

4 Conclusion

This work presents a transfer learning–based methodology to address the significant challenge of high-quality labeled data scarcity in resistance spot welding (RSW) monitoring. Such a limitation is typical of industrial environments, where the destructive tests required for reliable data labeling are expensive and time-intensive. An experimental campaign has been conducted to obtain the target domain dataset D_t, which contains 48 welding points made under various process parameters. Following feature extraction, a neural network has been optimized to predict the diameter of the weld nugget, achieving a 7.71% MAPE after the feature selection and hyperparameter tuning processes. Consequently, a model based on 108 welds made under different experimental conditions (materials thikness, parameters, electrodes diameter) that predicts the TST peak load has been used to construct three TL-based models, namely the *FE*, *FT*, and $FF - T$ to be trained on the target domain dataset, predict the nugget diameter, and be compared with the D_t model performances. Following, the best MAPE has been reduced by 25%, being 5.76% for the *FE* model. Notably, the *FE* model shows more stable convergence behavior compared to training from scratch, suggesting potential for improved generalization when source domain knowledge is properly leveraged. Consequently, it has been shown that the source learning task of predicting the peak load is an ideal one for the target learning task of weld nugget diameter prediction.

To conclude, source domain datasets can be utilized for training models whose knowledge can be subsequently transferred to a different problem or setup. Transfer learning methods could serve as an effective tool to address the common challenge of limited and imbalanced production line data. Further investigation is required, and future research may investigate the generalizability of this approach across various welding machines and guns, as well as its applicability when transitioning from controlled laboratory settings to real-world production environments.

Acknowledgements This study was supported by J-Tech@PoliTo, advanced joining technologies research center at Politecnico di Torino. The authors thank Mr Matteo Perrone for his help and assistance during the experimental campaign.

Author Contributions All the authors contributed equally to the work.

Funding Open access funding provided by Politecnico di Torino within the CRUI-CARE Agreement.

Declarations

Competing interests Not applicable.

References

1. Zhao E, Meng J, Lu M, Li X, Tong T, Fu X, Liu P, Zhang J, Zhao H (2025) Mechanical property curve of resistance spot welding for on-line monitoring of the welding quality. J Manuf Process 150:539–554. https://doi.org/10.1016/j.jmapro.2025.06.061
2. Zhao D, Vdonin N, Slobodyan M, Butsykin S, Kiselev A, Gordynets A, Wang Y (2024) Dynamic resistance signal–based wear monitoring of resistance spot welding electrodes. Int J Adv Manuf Technol 133(7–8):3267–3281. https://doi.org/10.1007/s00170-024-13993-y
3. Xia YJ, Zhou L, Shen Y, Wegner DM, Haselhuhn AS, Li YB, Carlson BE (2021) Online measurement of weld penetration in robotic resistance spot welding using electrode displacement signals. Measurement: Journal of the International Measurement Confederation 168. https://doi.org/10.1016/j.measurement.2020.108397
4. Russell M, Kershaw J, Xia Y, Lv T, Li Y, Ghassemi-Armaki H, Carlson BE, Wang P (2023) Comparison and explanation of data-driven modeling for weld quality prediction in resistance spot welding. J Intell Manuf. https://doi.org/10.1007/s10845-023-02108-1
5. Mathiszik C, Koal J, Zschetzsche J, Füssel U, Schmale HC (2024) Study on precise weld diameter validations by comparing destructive testing methods in resistance spot welding. Welding in the World 68(7):1825–1835. https://doi.org/10.1007/s40194-024-01747-z
6. Moghanizadeh A (2016) Evaluation of the physical properties of spot welding using ultrasonic testing. Int J Adv Manuf Technol 85(1–4):535–545. https://doi.org/10.1007/s00170-015-7952-y
7. Santoro L, Razza V, De Maddis M (2024) Frequency-based analysis of active laser thermography for spot weld quality assessment. Int J Adv Manuf Technol 130(5–6):3017–3029. https://doi.org/10.1007/s00170-023-12845-5
8. Vijayan V, Murugan S, Son SG, Park YD (2020) Microstructural analysis of cavity formed in advanced high-strength steel resistance spot welds. J Mater Eng Perform 29(10):6372–6377. https://doi.org/10.1007/s11665-020-05139-4
9. Mathiszik C, Zschetzsche E, Reinke A, Koal J, Zschetzsche J, Füssel U (2022) Magnetic characterization of the nugget microstructure at resistance spot welding. Crystals 12(11). https://doi.org/10.3390/cryst12111512
10. Podržaj P, Polajnar I, Diaci J, Kari Z (2004) Expulsion detection system for resistance spot welding based on a neural network. Meas Sci Technol 15(3):592–598. https://doi.org/10.1088/0957-0233/15/3/011
11. Kershaw J, Ghassemi-Armaki H, Carlson BE, Wang P (2024) Advanced process characterization and machine learning-based correlations between interdiffusion layer and expulsion in spot welding. J Manuf Process 109:222–234. https://doi.org/10.1016/j.jmapro.2023.12.013
12. Xia YJ, Song Q, Yi B, Lyu T, Sun Z, Li Y (2025) Improving out-of-distribution generalization for online weld expulsion inspection using physics-informed neural networks. Welding in the World 69(5):1309–1322. https://doi.org/10.1007/s40194-025-01950-6
13. Bogaerts L, Dejans A, Faes MGR, Moens D (2023) A machine learning approach for efficient and robust resistance spot welding monitoring. Welding in the World 67(8):1923–1935. https://doi.org/10.1007/s40194-023-01519-1
14. Wang PE, Ghassemi-Armaki H, Pour M, Zhao X, Ma J, Sattari K, Carlson B (2025) Applicable and generalizable machine learning for intelligent welding in automotive manufacturing. Welding in the World 69(5):1349–1384. https://doi.org/10.1007/s40194-025-01951-5
15. Von Rueden L, Mayer S, Beckh K, Georgiev B, Giesselbach S, Heese R, Kirsch B, Pfrommer J, Pick A, Ramamurthy R, Walczak M, Garcke J, Bauckhage C, Schuecker J (2023) Informed machine learning - a taxonomy and survey of integrating prior knowledge into learning systems. IEEE Trans Knowl Data Eng 35(1):614–633. https://doi.org/10.1109/TKDE.2021.3079836
16. Russell M, Wang P (2023) Maximizing model generalization for machine condition monitoring with self-supervised learning and federated learning. J Manuf Syst 71:274–285. https://doi.org/10.1016/j.jmsy.2023.09.008
17. Azari MS, Flammini F, Santini S, Caporuscio M (2023) A systematic literature review on transfer learning for predictive maintenance in industry 4.0. IEEE Access 11:12887–12910. https://doi.org/10.1109/ACCESS.2023.3239784
18. Guo P, Zhu Q, Kang J, Wang Y, Hu W (2022) Quality assessment of rsw based on transfer learning and imbalanced multi-class classification algorithm. IEEE Access 10:113619–113630. https://doi.org/10.1109/ACCESS.2022.3212410
19. Noh I, Jeon Y, Lee SW (2023) Development of robust fault diagnosis model for variable situations in robotic spot-welding (rsw) process based on transfer learning. J Mech Sci Technol 37(12):6123–6129. https://doi.org/10.1007/s12206-023-2405-2
20. Xiao M, Yang B, Wang S, Chang Y, Li S, Yi G (2023) Research on recognition methods of spot-welding surface appearances based on transfer learning and a lightweight high-precision convolutional neural network. J Intell Manuf 34(5):2153–2170. https://doi.org/10.1007/s10845-022-01909-0

21. Santoro L, Razza V, De Maddis M (2024) Nugget and corona bond size measurement through active thermography and transfer learning model. Int J Adv Manuf Technol 133(11–12):5883–5896. https://doi.org/10.1007/s00170-024-14096-4
22. Yue ZJ, Chen QR, Bao ZG, Huang L, Tan GB, Hou ZH, Li MS, Huang SY, Zhao HL, Kong JY, Wang J, Liu Q (2024) Improving rsw nugget diameter prediction method: unleashing the power of multi-fidelity neural networks and transfer learning. Adv Manuf 12(3):409–427. https://doi.org/10.1007/s40436-024-00503-2
23. Giannetti C, Essien A (2022) Towards scalable and reusable predictive models for cyber twins in manufacturing systems. J Intell Manuf 33(2):441–455. https://doi.org/10.1007/s10845-021-01804-0
24. BS EN ISO 17677-1-2021 - Resistance welding (2021) Vocabulary. Part 1, Spot, projection and seam welding
25. Nielsen KL, Tvergaard V (2010) Ductile shear failure or plug failure of spot welds modelled by modified gurson model. Eng Fract Mech 77(7):1031–1047. https://doi.org/10.1016/j.engfracmech.2010.02.031
26. Pouranvari M, Asgari H, Mosavizadch S, Marashi P, Goodarzi M (2007) Effect of weld nugget size on overload failure mode of resistance spot welds. Sci Technol Weld Joining 12(3):217–225. https://doi.org/10.1179/174329307X164409
27. BS EN ISO 14273–2016-Resistance welding (2016) Destructive testing of welds. Specimen dimensions and procedure for tensile shear testing resistance spot and embossed projection welds
28. BS EN ISO 14373–2015-Resistance welding (2015) Procedure for spot welding of uncoated and coated low carbon steels
29. Zhang H, Senkara J (2011) Resistance Welding - Fundamentals and Applications, 2nd edn. https://doi.org/10.1201/b11752
30. Ma Y, Wu P, Xuan C, Zhang Y, Su H (2013) Review on Techniques for On-Line Monitoring of Resistance Spot Welding Process. Adv Mater Sci Eng 2013:1–6. https://doi.org/10.1155/2013/630984
31. Chandrashekar G, Sahin F (2014) A survey on feature selection methods. Comput Electr Eng 40(1):16–28. https://doi.org/10.1016/j.compeleceng.2013.11.024
32. Pirbazari AM, Chakravorty A, Rong C (2019) Evaluating feature selection methods for short-term load forecasting. In: 2019 IEEE International Conference on Big Data and Smart Computing, BigComp 2019 - Proceedings. https://doi.org/10.1109/BIGCOMP.2019.8679188
33. Ahmadi A, Daccache A, Snyder RL, Suvočarev K (2022) Meteorological driving forces of reference evapotranspiration and their trends in california. Sci Total Environ 849. https://doi.org/10.1016/j.scitotenv.2022.157823
34. Xiang F, Zhao Y, Zhang M, Zuo Y, Zou X, Tao F (2024) Ensemble learning-based stability improvement method for feature selection towards performance prediction. J Manuf Syst 74:55–67. https://doi.org/10.1016/j.jmsy.2024.03.001
35. Feng S, Zhou H, Dong H (2019) Using deep neural network with small dataset to predict material defects. Mater Des 162:300–310. https://doi.org/10.1016/j.matdes.2018.11.060
36. Panza L, De Maddis M, Russo Spena P (2022) Use of electrode displacement signals for electrode degradation assessment in resistance spot welding. J Manuf Proc 76:93–105. https://doi.org/10.1016/j.jmapro.2022.01.060
37. Panza L, Bruno G, Antal G, De Maddis M, Russo Spena P (2024) Machine learning tool for the prediction of electrode wear effect on the quality of resistance spot welds. Int J Interac Design Manuf (IJIDeM). https://doi.org/10.1007/s12008-023-01733-7
38. Antal G, Razza V, De Maddis M (2025) Frequency domain signal analysis based index for expulsion quantification in resistance spot welding. Int J Adv Manuf Technol 136(3):1149–1161. https://doi.org/10.1007/s00170-024-14922-9
39. Zhao D, Wang Y, Zhang P, Liang D (2019) Modeling and experimental research on resistance spot welded joints for dual-phase steel. Materials 12(7). https://doi.org/10.3390/ma12071108