A machine learning approach for efficient and robust resistance spot welding monitoring

Lars Bogaerts[1*], Arnout Dejans[1], Matthias Faes[2] and David Moens[1]

[1*]Department of Mechanical Engineering – LMSD division, KU Leuven, Jan De Nayerlaan 5, St.-Katelijne-Waver, 2860, Belgium.
[2]Chair for Reliability Engineering, TU Dortmund University, Leonhard-Euler-Strasse 5, Dortmund, 44227, Germany.

*Corresponding author(s). E-mail(s): lars.bogaerts[at]kuleuven.be;
Contributing authors: Arnout.dejans[at]kuleuven.be; matthias.faes[at]tu-dortmund.de; david.moens[at]kuleuven.be;

Abstract

The estimation of the weld nugget diameter generated by the resistance spot welding process is a crucial element in the assessment of the overall quality of the weld and plays a major role in in-line process control. The process is crucial to produce end-products in many industries such as aviation, aerospace, automotive and other industrial areas. A modern car body contains typically several thousands of welds produced by resistance spot welding, setting an ideal scene for in-line process control. Current state of the art monitoring methods are based on several features extracted from the dynamic resistance signal. However, the accuracy of those is generally not high. In this work, a method for predicting the nugget diameter based on the combination of unsupervised deep learning and Gaussian process regression is developed. Autoencoders are adopted to extract features from the dynamic resistance curve in a low dimensional representation. These features embodies underlying information on the process, possibly unobservable or not detectable by any other currently existing approach. Next, a Gaussian process regression model is trained to link those features to the target weld nugget diameter. Compared with the currently popular geometrical attributes approach, the results show that the model has a higher prediction accuracy in nugget diameter prediction, whilst remaining a low cost implementation in an industrial setting. These results are supported by several cases, derived directly from common industrial bottlenecks. Both cases indicate a strong potential with the new AE-GPR approach, with consistently improved results compared to the currently popular geometrical attributes approach.

Keywords: Resistance spot welding, Nugget diameter, deep learning, machine learning

1 Introduction

Resistance spot welding (RSW) is a highly efficient, low cost and easy realisable joining technique. The technique has been used extensively and is crucial to produce end-products in many industries such as aviation, aerospace, automotive and other industrial areas. The process employs two welding electrodes that press two or more overlapping sheet-like workpieces together. The heat generated by the then-applied electric current, in correspondence with Joule's law, causes local melting at the workpiece's common faying surface, leading to joining these workpieces. This ease of operation contributed vastly to the quality and automation of the production of modern car bodies, which contain typically several thousands of welds produced by RSW. The safety and reliability of current automobile industry is only one of many examples that notably profited from the valuable RSW process.

To a large extent, this success can be attributed to the ease of automation of the RSW process in an assembly line. Further, RSW also combines high strength joining with production flexibility, low cost and fast throughput. Thanks to its wide-spread use, RSW has grown to a mature joining technique, with literature dating back to the second half of the 19th century (experimentally oriented) [8] and early 2000's (on-line monitoring, Finite Element methods) [12, 13]. Yet, in industrial practice, it still suffers from a high sensitivity to often uncontrollable and variable process conditions. This makes the RSW process extremely vulnerable to environmental effects, surface conditions, misalignment, wear, etc. Inevitably, abnormal welding conditions dramatically reduce the consistency of welds, which generally leads to significant degradation of the weld quality. Therefore, it is a vital task to control and monitor the quality of the welding process [10]. Also, in order to increase productivity and achieve a robust final assembly, an attempt to minimise the number of required spot welds is made. This is only possible when consistent and sufficient weld quality can be guaranteed [23, 32]. For the latter, a common technique remains non-destructive testing, based on a random subset of the workpieces on the production site. However, these weld quality estimations can only be examined off-line, making it impossible to receive brisk and pertinent information. Furthermore, it is very cost inefficient for mass production environments, where RSW is vastly present.

In the context of process monitoring, real-time weld quality estimation based on data-driven techniques are becoming ever more common [3, 11, 21]. These approaches typically link process parameters and on-line measurements to product quality metrics in order to guard the process. In this respect, machine learning approaches yield very fast black-box models enabling on-line application for process control. However, multiple problems arise with these black-box models: (a) These models are known to have only limited value when extrapolation is required. Their use is most relevant for well-confined and -controlled processes that enable the generation of a clear, industrially representative and comprehensive data set for the model training. There are large discrepancies between a lab-environment or numerically-made data-set and an industrial data-set, which is prone to inaccuracy due to changing variables or boundary conditions that come with the large change in environment. (b) Quantitative measurement capabilities for the process response are limited in the RSW process, hence eliminating the industrial applicability of data-hungry algorithms such as most supervised deep-learning toolboxes [20].

Various in-line measurement techniques for the RSW process are investigated in literature, and can be classified based on the quantity of measurement (e.g., force, current, time) and their corresponding measurement device. Some of these techniques show promising results regarding in-line prediction of the weld nugget diameter, which is usually the primary choice for the Quality Indicator (QI) of the process [4]. A first class of prediction models makes use of mechanical measurements, e.g., displacements [24], forces [31] or acoustic emission [7]. While these are possible sources of valuable process information, and state of the art technology for measurement of the required quantities is proven achievable on industrial scale, the reliability, accuracy and flexibility of these models remain a challenge [25], leading to their limited use in an industrial context. A second class of techniques focuses on monitoring through electrical signals. In this context, dynamic resistance (DR) measurements are widely investigated and implemented in industrial practice [9, 22, 29]. There are several milestones in the progress of monitoring the dynamic resistance. In 2002, Cho and Rhee [5] calculated dynamic resistance based on current and voltage from the primary part of the transformer. Further breakthroughs include quality estimators by means of a Hopfield network, presented in [6], Artificial Neural Networks (ANN) [17], welding quality classifiers by means of Probabilistic Neural Networks (PNN) [27] and a random forest model based on features of the dynamic resistance curves [25]. Measurement of dynamic resistance has become

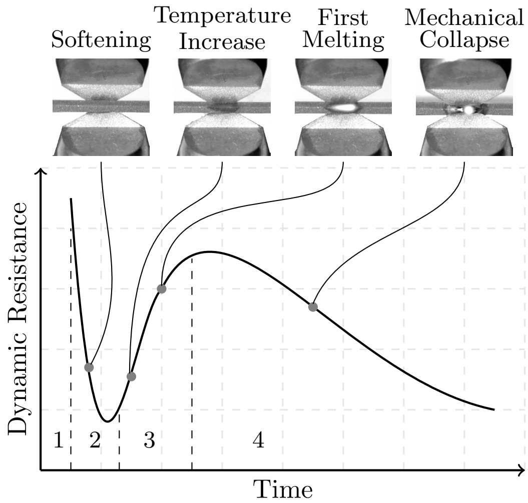

Fig. 1: Theoretical dynamic resistance curve interpretation and characterisation. Top images illustrate the evolution of the weld nugget at the given phase and are made by a phantom VEO 640 ultra high speed camera. Based on the original graph of D.W. Dickinson et al. [8]

the accepted paradigm in industry [1, 15, 30], and has been implemented in several commercially available power sources as a quality monitoring and evaluation tool. For this reason, it is selected as primary feature in this work.Furthermore, there is a clear link between the evolution of the weld nugget and the dynamic resistance, as illustrated in figure 1. For more detail on the subject, the reader is referred to [8].

1.1 Motivation

While extensive literature exists regarding the online monitoring and the quality assessment of the spot welding process based on dynamic resistance and alternative measurements, many challenges remain. Many of the aforementioned techniques suffer from drawbacks that hinder their optimal cost-effective and fully automated application for RSW proceses in industrial practice. These shortcomings are the following.

1. Current dynamic resistance based techniques fail to establish an accurate prediction when variations on the input signal are present, either due to process parameter alterations or inherent randomness in the process.
2. The dynamic resistance curve is containing information that is not necessarily observable in a time signal, and therefore lost by currently existing techniques. Consequently, the widely investigated techniques based on the geometric attributes of the dynamic resistance measurement (see section 3.2) are not fully exploiting their potential.
3. Alternatives for the measurement of dynamic resistance are lacking robustness, mainly due to the type of measurement. This leads to either economical or infrastructural burdens, which render them less interesting for industrial application.

In an effort to remedy these shortcomings, this paper investigates a weld quality monitoring approach based on the underlying parametric dependencies of the dynamic resistance during the RSW process. The weld nugget diameter serves as the main driving Quality Indicator (QI) for this research. The capability of predicting the weld nugget diameter by means of limited measurements is an important aspect of the developed approach. This is important to keep the number of man-hours required to measure the data-set feasible. Furthermore, the technique should be able to cope with variations, re-calibrations or other possible variability within industrial application.

We propose deep learning autoencoders to discover a low dimensional representation that captures the underlying causes of the resistance in the secondary circuit of the welding machine during the RSW process. This allows for a sparse representation that can be leveraged towards inline prediction of the weld nugget diameter. The method is demonstrated on an experimental dataset to clarify the technique and to compare it with the geometrical attributes approach (elaborated in section 3), by means of computational performance, prediction accuracy, advantages and disadvantages.

Next, the method is demonstrated on several example problems, showing that the technique has large potential on diverse problems.

The paper is structured as follows:

- Section 2 elaborates on the conducted experiments, as well as the required hardware both for experiments, measurements and processing the data,

- Section 3 discusses the geometrical attributes approach, which serves as a reference for the novel methods developed in the next sections,
- Section 4 introduces the autoencoder based approach,
- Section 5 applies the introduced method to two case studies to illustrate its application and performance,
- Section 6 lists the most important conclusions of this manuscript.

2 Experimental setup

Experiments described in this work are performed on an ARO servo-driven RSW machine with a 1000 Hz, 90 kVA DC power source, as depicted in figure 2. The machine is equipped with water cooled electrode caps with an ISO 5821:2009 FE-15.8-5.5-30 geometry. Three experiments were conducted using process parameters as summarised in table 1. For each experiment, the complete set of process parameters and the number of realisations n is provided. Experimental set 1 represents a limited run in a production setup under ideal conditions, where variations are only generated due to randomness in the process and systematic uncertainty due to measurement accuracy. Experimental set 2 is a set of multiple experiments, with each subset having a different value for a selected process parameter. For this work, the current is adjusted, which is one of the most influential process parameters of the process [8]. The rationale behind this case study is to prove the flexibility of the technique over a wide range of machine parameters. Experimental set 3 is a set of multiple experiments, with each subset having a geometrical adjustment of the electrode clamp, causing a variation in the static electrode bulk resistance (EBR), which is part of the secondary circuit. This is realised by adding custom build raisers *(shims)* between the electrode and its holder. Due to the principle of stacking multiple thin shims, the conductivity decreases significantly, causing the resistivity to increase with only a minor increase in extra material required. For the three examined cases, the resistance between two predetermined points, one above the added shims and one below the added shims, is measured to be 11.9, 108.4 and 371.5$\mu\Omega$ respectively. This case is included to illustrate the performance of the technique for a common problem in an industrial setting, where the change of welding electrodes or welding clamps inherently causes a variation in the overall resistance of the machine's electrical circuit. The welded specimen are low carbon steel samples of 20 by 70 mm and thickness of 1 mm and are welded in as-delivered condition. Data is acquired by a Dewetron DEWE2-a4L data acquisition system at a sampling rate of 2 MHz. Acquired signals include (1) electrical voltage over the welding electrodes and (2) electrical current in the welding circuit, measured by a PEM RFT 300S Rogowski coil and preamplifier.

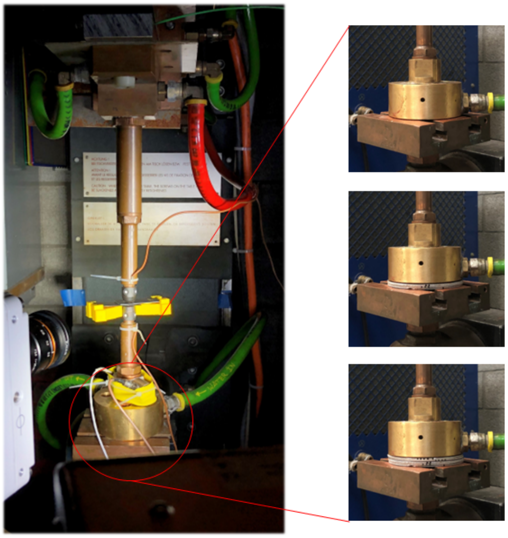

Fig. 2: Setup of the machine *(left)*, with location of shims for experimental set 3 *(right)*, top-down: zero, 3 and 5 shims.

All samples are labelled by physical measurements of the weld nugget diameters according to ISO 10447:2015 *(specifies the procedures and recommended tooling to be used for peel and chisel testing of resistance spot and projection welds. ISO 10447:2015 applies to welds made in two or more sheets in the thickness range of 0,5 mm to 3,0 mm)*. The process parameters are chosen such that the nominal nugget diameters correspond to the welding lobe diagram according to ISO 14327:2004, $3.5\sqrt{t}$, with t the thickness of a single plate. Furthermore, the weld time and force differ greatly between the conducted experiments.

This is deliberately determined as such, to provide a demonstration on a case where the weld is generated in a short time window as well as a case where a weld nugget is formed slower.

3 Dynamic resistance based monitoring

This section describes, discusses and illustrates the geometrical attributes regression model as presented in literature. It serves as a reference for the developments in the remainder of this work.

The geometrical attributes regression model is currently the most commonly adopted prediction model in industry. The model is based on input-output pairs, respectively from on-line measurements and nugget diameter measurements stemming from destructive testing (as elaborated in section 2). Figure 3 visualises the main flow of the approach. It consists of k welds being generated and measured by means of an experimental campaign. The samples are *peeled*, a destructive testing method (ISO 10447:2015) for determining the diameter of the weld nugget. When the demanded samples k are generated, the data is subjected to a training algorithm based on the input-output pairs. All additional samples ($i \geq k$) are then predicted based on the trained model.

3.1 Feature extraction

Measurement of dynamic resistance is one of the most effective techniques for quality monitoring and estimation, aided by the fact that measurements are straightforward, including electrical current and voltage in the secondary circuit. Next, the dynamic resistance signal is obtained according to Ohm's law, i.e.

$$R(t) = \frac{U(t)}{I(t)}, \tag{1}$$

with $R(t)$ the dynamic resistance, $U(t)$ the the voltage, $I(t)$ the welding current and t the welding time.

During the process, the welding machine forms a closed circuit with the secondary circuit of a transformer, ensuring a solid mechanical assembly between tooling and work pieces. The closed circuit is modelled in terms of their individual resistances. In this resistance model, the electrical resistances of the transformer, the mechanical assembly, and the work pieces are represented as respectively R_t, R_m and R_l. The resistances R_t and R_m are assumed constant during the process. The resistance of the work pieces is split into three components:

1. the bulk resistance of the sheet metal (R_b),
2. the interface resistance between electrodes and sheet metal (R_c),
3. the contact resistance of the parts surfaces (R_f).

For the general case of two pieces of sheet metal, assumed equal in properties and dimensions, the resistance of the work pieces is:

$$R_l = 2R_b + 2R_c + R_f. \tag{2}$$

As illustrated in figure 1, there are multiple stages during the process, causing a large fluctuation of R_l. The geometrical attributes approach,

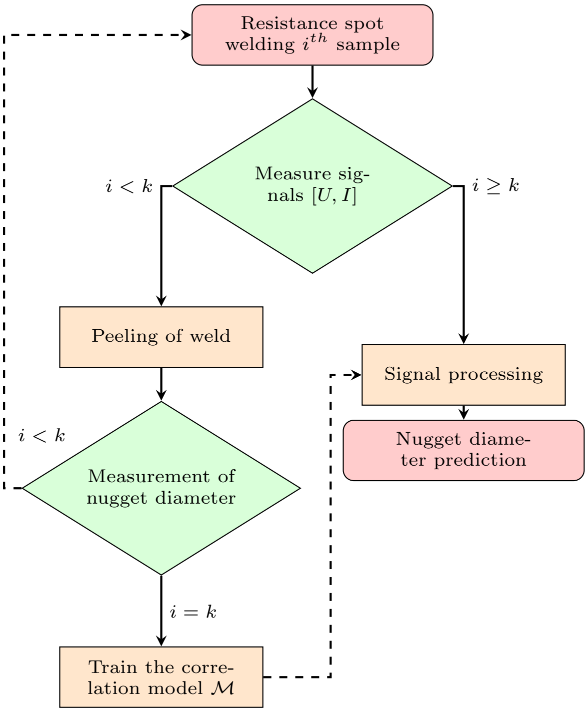

Fig. 3: Workflow a. General prediction model structure for RSW

Table 1: Overview of conducted experiments and their process parameters used in this paper. *EBR: Electrode bulk resistance setup, n: amount of welds*

No.	Current (kA)	Force (kN)	Time (ms)	EBR	N
1	7.2	5	210	cte	174
2	6.0 - 6.6 - 7.2	2.5	70	cte	50 - 50 - 50
3	7.6	2.5	80	var.	50 - 50 - 50

elaborated in section 3.2, refers to several points that characterise the dynamical resistance curve. Extracted points from the curve are based on e.g., peaks and slopes. Figure 4 illustrates several key points of the curve, respectively the initial peak R_0 in phase 1, the pit R_α in phase 2, the peak at the beginning of phase 4, also commonly known as the beta peak R_β and the last value of the DR curve R_γ, also respectively the times t_0 t_α, t_β and t_γ. Next, several critical derivatives are selected, the mean value R_m, the slopes m_1 and m_2 and the resistance variance $dR1 = R_\beta - R_\alpha$ and $dR2 = R_\gamma - R_\beta$

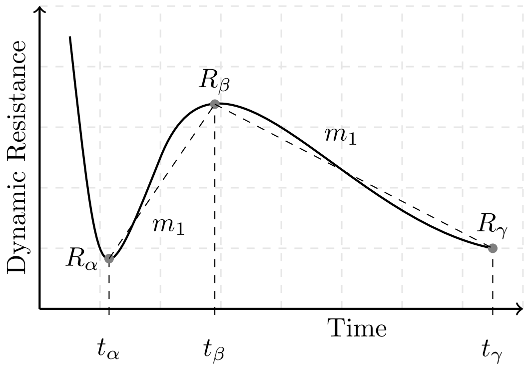

Fig. 4: Theoretical dynamic resistance curve, with selected features for the geometrical attributes approach, adopted from [9].

However, one drawback of this technique is that these measured parameters fluctuate heavily during a single weld due to the time-varying current generated by a mid frequency direct current (MFDC) power supply. The dynamic resistance, derived according to Eq. 1, is subjected to a signal filter in order to obtain the main trend. *Zhang et al.* [26] evaluated the raw signals and acknowledged, based on Fourier spectrum analysis, that periodic features are key to the large fluctuations. They applied a fourth-order digital low pass filter with a cut-off frequency of 50 Hz. For this study, the concept of the moving average filtering is selected to eliminate the interference of periodic signals effectively [33].

Table 2: Overview of correlations between derived features and the QI, based on experimental case 1

feature (f_i)	%	feature (f_i)	%
t_0	1.68	R_0	3.06
t_α	-31.37	R_α	33.36
t_β	-30.60	R_β	20.62
t_γ	-20.43	R_γ	-8.56
m_1	8.60	m_2	6.30
dR_1	-4.45	dR_2	49.73
R_m	44.29		

3.2 Geometrical attributes model

Post-filtering the dynamic resistance signal provides a noise-free curve where the features described in section 3.1 (Figure 4) are observable. Table 2 gives an overview of the correlations between the features according to figure 4 and the QI, based on case 1 from table 1. This table shows that correlations between inputs and the QI are present. This confirms the applicability of the geometrical attributes approach, as discussed by [9]. The ensemble of these, or similarly derived, correlations, superseded with generic regression analysis, are current state of the art methods for RSW quality indication. This is visualised as the left part in figure 5. Part A, the feature selection, yields the aforementioned features of the dynamic resistance curve. Next, in part B, a multiple linear regression model describes the relationship between the features and the QI. The selection of features is case dependent, and relates to the most significant features of the curve, with a minimum of six points, without significantly affecting the model performance [9]. The regression model is described as

$$q = \beta_0 + \beta_1 f_1 + \beta_2 f_2 + ... + \beta_k f_k + \varepsilon \quad (3)$$

with q the QI response, f the regressor variable, β the regression coefficients and ε the error term.

This overall workflow makes it possible to project new data-points onto the regression model and predict (interpolate) an estimate for the QI.

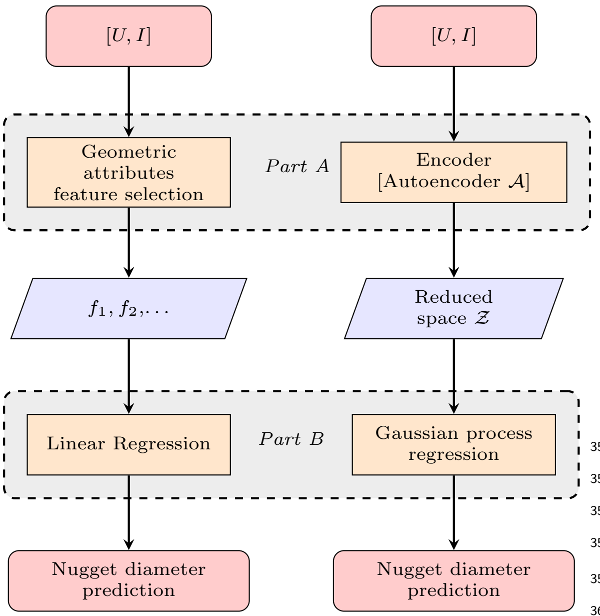

Fig. 5: Workflow b. Technical

4 Autoencoder based weld quality monitoring

This section discusses the main innovation of this study, replacing both part A and part B of the technique elaborated in section 3.2, as also illustrated in the right part of figure 5. In the following, the method is introduced and discussed, as well as illustrated based on experimental set 1 (table 1).

4.1 Part A: Feature extraction

4.1.1 The curse of dimensionality

Part A in this work focuses on improving the amount of information that is extracted from the dynamic resistance curve. For this, an efficient coding is required that is capable of learning a low dimensional representation for a set of data. Opposed to the method of manually selecting points, recent advances in deep learning methodologies are proven to be very efficient in gathering interesting features in the data, which are possibly unobservable or not detectable to the engineer during manual evaluation of the data. However, especially for this application, this poses specific challenges regarding the architecture of the network. Indeed, with a sampling rate of $2Mhz$ and the process yielding $> 200ms$ of welding time, a very high amount of data is gathered on multiple channels, which serves as the input of the network. The main problem is that the computational cost of the neural network scales exponentially with the data, by cause of a connection that is required to each neuron in the next layer, according to:

$$z_j^l = \varphi(v_j) = \varphi\left(\sum_{i=0}^{k} w_i . z_i^{l-1}\right), \qquad (4)$$

where z_j^l is the value z for neuron number j in layer l, φ represents the activation function, usually a sigmoid function ranging from -1 to 1, $\varphi(v_j) = \frac{2}{(1+exp(-2.v_j))} - 1$, z_i^{l-1} representing neuron number i from layer $l-1$, w_i the weight assigned to each connection with the previous layer, k the number of neurons in layer $l-1$ and $z_0 = \pm 1$ for adding a bias $b = w_0$ to the summation operator, yielding v_j. Evaluating this key equation is computationally not a large effort. However, due to the architecture of neural networks, it has to be solved numerous times during training. As such the total training effort is increased drastically. This is often referred to as the curse of dimensionality, referring to problems that occur when dealing with data in high-dimensional spaces. It prevents strategies to work efficiently, while creating problems concerning computational expenses, which do not occur in low-dimensional spaces.

Multiple techniques for dimensionality reduction exist. They can be divided into convex and non-convex techniques, where convex techniques optimise an objective function that does not contain any local optima, e.g., Principal Component Analysis (PCA), Kernel PCA, Isomaps, Local Linear Embedding (LLE) and non-convex techniques optimise objective functions that do contain local optima, e.g., Locally Linear Coordination (LLC), manifold charting or autoencoders [14, 28]. For a competent model, capable of being deployed in an on-line context, a marginal computation cost is

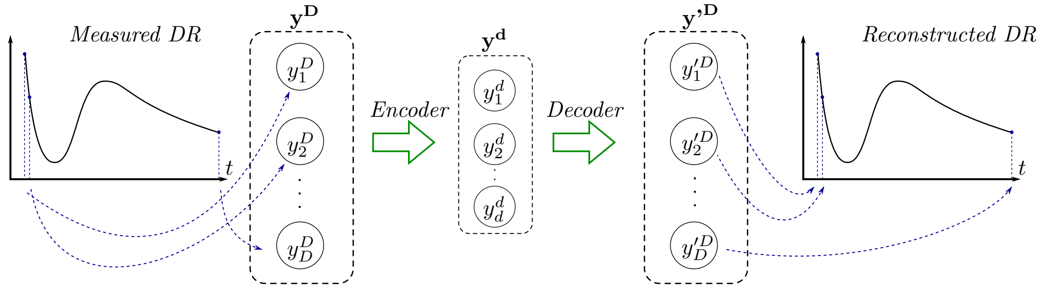

Fig. 6: Autoencoder topology

envisaged for the projection of a series of points to the low dimensional space, but also the ability to embed new high-dimensional data points into an existing low-dimensional data representation is important. For these reasons, autoencoders have been selected for the dimension reduction in this work.

4.1.2 Autoencoder based dimension reduction

Autoencoders, a type of artificial neural network, are the proposed solution for this work, based on their efficiency, non-linear transformation and intuitive nature [2].

These feed-forward networks have an odd number of hidden layers hL_i, with $i = 1 \ldots n_l$ and n_l the amount of layers, where the hidden layers are dimensioned such that the layer in the middle has a lower amount of neurons than the first and last layer. This separates the autoencoder in an input layer, an encoder part, the middle layer with $d \leq D$, a decoder part and the reconstructed layer:

$$\mathbf{y^D} \overset{Encoder}{\rightarrow} \mathbf{y^d} \overset{Decoder}{\rightarrow} \mathbf{y'^D}, \tag{5}$$

where $\mathbf{y^D}$ is the measured data, D the number of time-steps in the input data, and d the amount of neurons in the middle layer. The objective of the autoencoder is to generate this neural network architecture such that $\mathbf{y'^D} \approx \mathbf{y^D}$. The autoencoder is an unsupervised learning technique, since its goal is to minimise an error in reconstructing $\mathbf{y^D}$. The input layer $\mathbf{y^D}$ has D neurons, where each neuron represents an individual parameter from the dataset. This data is reconstructed in the final layer $\mathbf{y'^D}$ of equal dimension D, as illustrated in figure 6.

The centre layer in the network $\mathbf{y^d}$ represents the original data in a lower dimension d, while preserving as much structure as possible from the dataset $\mathbf{y^D}$. The resulting low dimensional representation in this centre layer functions as the input for further processing, which has the benefit of working with far less data without losing essential information.

Mapping from the input vector to another vector by means of an encoder, based on the general equation of a neural network topology (eq 4) gives:

$$y^d = \varphi(W^1 y^D + b^1) \tag{6}$$

and for the reconstruction through a decoder:

$$y'^D = \varphi(W^2 y^d + b^2). \tag{7}$$

The network is trained by minimising a loss function, which includes regularisation terms. Apart from the mean squared error function, an L2 regularisation term $\lambda * \Omega_{weights}$ and sparsity regularisation term $\beta * \Omega_{sparsity}$ are added to the loss function. The L2 regularisation term forces the weights to remain small, by adding a penalty to the loss function when weights are increasing. The sparsity regularisation term attempts to enforce a constraint on the sparsity of the output from the hidden layer. The cost function for training the autoencoder based on N samples yields

$$\mathcal{L}(y^D, y'^D) = \frac{1}{N} \sum_{n=1}^{N} \sum_{j=1}^{D} \left(y^D_{jn} - y'^D_{jn}\right)^2 + \lambda * \Omega_{weights} + \beta * \Omega_{sparsity} \tag{8}$$

with λ the coefficient for the L2 regularization term and β the coefficient for the sparsity regularization term.

The workflow for applying this metric for the dynamic resistance curve is illustrated in figure 7.

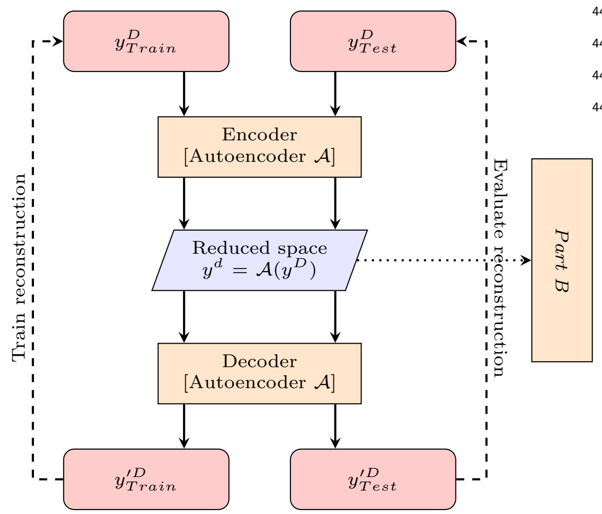

Fig. 7: Workflow c. Autoencoder training principle

In this figure, there is a clear distinction between the data for training and testing. It also indicates the main workflow, where eq. 8 is used to train the reconstruction of the measured signal and to evaluate the reconstruction of test data. Furthermore, the figure illustrates that part B is connected to the low dimensional layer, in the centre of the autoencoder.

4.1.3 Illustration

The autoencoder based dimension reduction is now illustrated based on experimental case 1 from table 1. Out of 174 experiments, 90 % serves as training data, whereas the remaining 10 % is test-data to evaluate the performance.

Figure 8 illustrates the performance of the network for one sample from y^D_{Test}, where one dynamic resistance curve is plotted next to its reconstructed counterpart. Since the error is nearly negligible, the instantaneous reconstruction error ϵ_{AE} is given, defined as:

$$\epsilon_{AE} = \frac{y'^D - y^D}{y^D}.100. \tag{9}$$

In this example, the signal is compressed from $D = 450.000$ data-points into a middle layer of the autoencoder, represented by y^d, where $d = 15$. The hyper-parameters for the network are summarised in table 3.

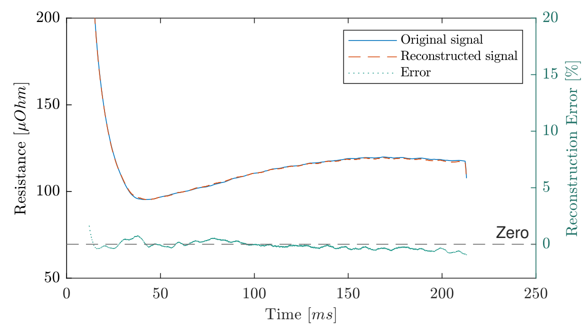

Fig. 8: Measured and reconstructed signal with means of an autoencoder network, with topology $450k - 15 - 450k$

Note that the reconstruction is only required for training of the autoencoder. For the purpose of dimensionality reduction, the encoder projection, resulting in the reduced space y^d is an important step to come to an efficient regression model, as illustrated in figure 7.

At this point, it can be concluded that the presented method is capable of projecting the measured dataset into a reduced space y^d, which acts as a low dimensional space. The reconstructed projected data, by means of the autoencoder, performs approximately equal to the measured data, as the error is nearly negligible. Furthermore, the projected data contains nearly all information to reconstruct the data in a low dimensional space, thus possessing at least as valuable information as the manually determined points, as described in section 3.2. Therefore, the reduced space y^d can serve as an input to current the QI prediction step using generalised techniques, e.g., multiple regression. It should be noted that the projected data has no physical meaning in the process, opposed to the selected features from section 3.2. The added benefit of this method is the robustness of

Table 3: Hyperparameters for training Autoencoder

Encoder-TF	*Sigmoid*
Decoder-TF	*Sigmoid*
d	15
L2 weight coef. [λ]	0.009
Sparsity Proportion [$\hat{\rho}_i$]	0.719
Sparsity Regularization [β]	1.08
Normalised data	*Yes*

the algorithm, which is unlikely to suffer abnormalites, e.g, where expulsion could cause an effect in the curve, resulting in misjudgement or wrong interpretation of the data. Additionally, the measured signal is prone to various effects like time shift, originated during the processing of the data, caused by the required filtering techniques of the signal. The presented approach is also insensitive to these effects.

4.2 Part B: QI prediction

Part B of the novel prediction model (see figure 5) requires an algorithm capable of giving a robust regression between input-output, respectively the measured signals during the welding process projected on a reduced space $\mathcal{Z}$ by means of an autoencoder, and the measured nugget diameter of the weld (QI). Due to the low amount of available labelled data, training a neural network based on input-output pairs tends to be troublesome, in particular overfitting is a main concern. This should not be confused with the neural network on which the autoencoder from section 4.1 is based, as the latter is an unsupervised technique, which does not require labelled data.

4.2.1 Gaussian Process Regression

A powerful tool, ideal for this problem is Gaussian Process Regression (GPR). In it's original form, Gaussian Process modelling is a statistical interpolation method that exploits Gaussian processes to interpolate a series of complex functions. The technique works well on small datasets, and has the capability to provide uncertainty metrics on the predictions. Gaussian process modelling, also known as Kriging, was introduced in the context of meta-modelling in the works of Sacks et al. [19], in which the original form of Kriging, as developed in the Master's thesis of D. Krige [16], served as a backbone to represent an input/output mapping of an expensive computational model. For application in machine learning, Kriging has evolved as both regression and classification tool, and has proven to be a treasured algorithm [18].

The required dataset $\mathcal{D}$ with N observation is presented as

$$\mathcal{D} = \left\{ \left(y_i^d, q_i\right) | i = 1, \ldots, N \right\} \tag{10}$$

with y_i^d the vector with multiple input variables and q_i as a measure for the QI, which is continuous, as this is a regression case.

In essence, GPR serves as an interpolation model for a regression problem mapping the QI on the variables.

They key equation of GPR is presented as:

$$q = \mathcal{M}(y) + \varepsilon, \tag{11}$$

with $\mathcal{M}(x)$ the GP model. It is assumed that the observed responses are noisy and the noise ε follows a zero-mean Gaussian distribution

$$\varepsilon \sim \mathcal{N}(0, \sum_n), \tag{12}$$

where $\sum_n$ is the covariance matrix of the noise term. For this work, the noise variance is identified as general heteroscedastic, in which the noise can differ for each observed response. For a more in depth study of GPR, the reader is referred to the citations in previous section.

An important feature of this type of model is the applicability of K-fold cross-validation. The main idea of this is to re-evaluate the model with a subset of the training data. This allows to optimise the hyper-parameters of the model, based on evaluating the model with each subset of data, thus minimising a cross-validation error. This yields a model, capable of working under limited data, and providing efficient predictors with minimal effort, unlikely to suffer from overfitting. Performance-wise, training this model takes significantly longer than linear regression. However, as elaborated in the introduction, the computational requirement for training is irrelevant for an in-line paradigm. What is of interest is the required time to evaluate a single sample. This will be studied for the illustrative data set in the next section.

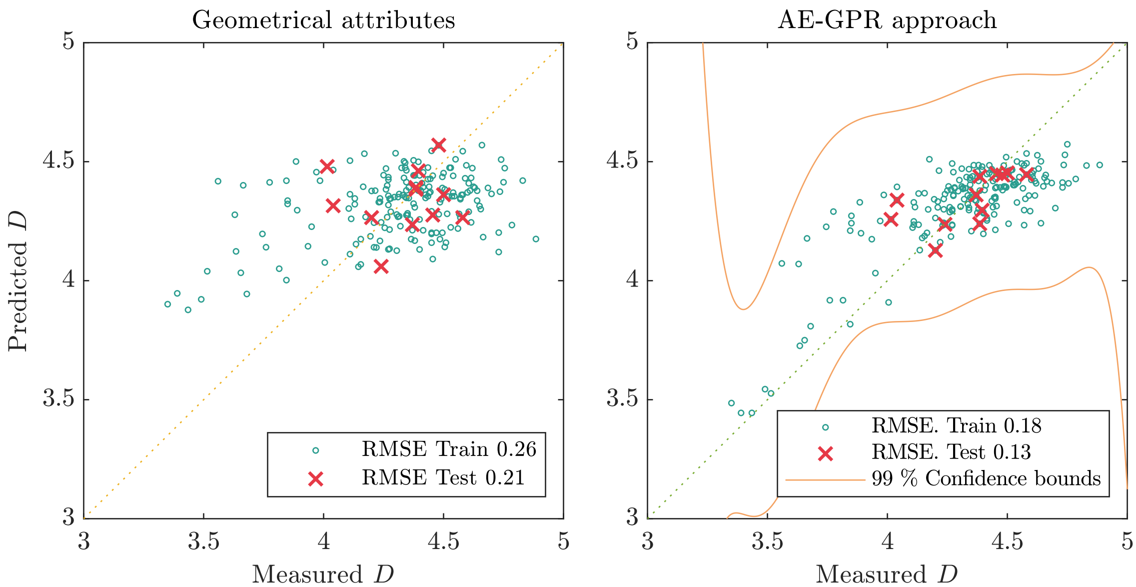

Fig. 9: Predicted vs. actual QI based on the AE-GPR approach *(right)* compared to the Geometrical attributes approach *(left)* for the reference case

4.2.2 Illustration

The methodology is presented on the data resulting from section 4.1.3. Computation experiments and the training of the proposed algorithms are performed on a server, consisting of 2 AMD Epyc 7601 32-core CPU's, 512 GB memory and an NVIDIA Tesla V100 - 32 GB graphical accelerator. For benchmarking computational effort of the proposed algorithm, the system is limited to only 1 core, while GPU acceleration is disabled.

The summary of computational costs for both the reference and the newly developed approach are given in table 4. Comparison of the required time for the regression approaches clearly reveals that the GPR technique is more expensive. This is due to the fact that substantially more time is required to determine the optimal hyperparameters, compared to the key, and only equation required for multiple linear regression 3. Apart from that, the table also shows that especially the training of the unsupervised deep learning based dimension reduction technique into an autoencoder (AE) necessitates a high performance computational device. However, the cost for training is only a small investment for a continuously guarded process. Table 4 also summarises the computational costs for training and propagating a single sample. This shows that propagating a single experiment through the full AE-GPR framework requires only a marginally higher cost compared to the approach based on the geometrical attributes.

Table 4: Computational costs for the framework, *AE: Autoencoder based dimension reduction, GPR: Gaussian Process Regression*

		Training	1 Sample
Part A	Geom. attributes	1.8 x 10^3 s	0.46 s
	AE	5 x 10^4 s	0.83 s
Part B	Linear regression	1.01 s	0.03 s
	GPR	91 s	0.04 s

In order to assess the performance of the developed method, figure 9 illustrates the correlation between the predicted QI and the measured QI, both for the geometrical attributes approach (*left*) and the novel AE-GPR approach *(right)*.

A clear improvement over the geometrical approach is noticeable. Not only does the Root Mean Squared Error (RMSE) improve vastly, from 0.26 to 0.18, it is also clearly visible that the geometrical approach fails to predict the overall nugget diameter. Specifically, the graph based on the geometrical approach shows a horizontal scatter, indicating a larger spread on the measured QI, but a small spread on the predicted

QI. Furthermore, in addition to the continuous mean regressor estimate by the GPR model, also confidence bounds are provided, based on the available training data and known variance of the GPR model. For this case, the 99 % bounds are provided, showing a narrow band encapsulating the data. As a final verification, the algorithm is subjected to several testing points, which were excluded from the training. The latter is represented by 10 % of the available data and is visualised by the red crosses. The RMSE for the test points improves from 0.21 to 0.13 for the test data, which emphasises the performance of the model by resulting in a prediction well within the bounds predicted by the algorithm.

5 Verification of the proposed method

To benchmark the accuracy and robustness of the introduced approach, the algorithm is subjected to two more cases. Compared to case 1 of section 3.2 for which an experimental set with fixed process parameters was used (thus only yielding process variability on both input and output), the next two cases include highly relevant industrial events representing common process variations.

5.1 Current variation

This case (table 1, No.2) is a dataset combining data from three consecutive experimental runs with respectively altering the current as described in table 1. The goal of this case is to determine the robustness of the AE-GPR approach based on a broader set of input signals, as could be the case in an industrial setup where low- and flexible production is key and in-line monitoring is envisaged. First, applying a dimension reduction on the measured signals reduces the dataset from 450.000 samples to 15, while only losing < 1 % of information, based on mean-squared error estimates. This is illustrated in figure 11, where the original curve, the reconstructed curve and the error are illustrated. Analysing the measured and predicted data, illustrated in figure 10, clearly reveals several meaningful results.

First, no separate data clusters can be distinguished according to the difference in welding currents. Yet, the three datasets are separated by colour, which reveals the difference in nugget diameter in function of process parameters. Second, the measured results with $6.0kA$ and $7.2kA$ have a wider spread of data compared to the set welded with $6.6kA$. Third, advancing to the results of the prediction model, there is a clear improvement of the AE-GPR method over the geometrical approach. The RMSE is improved , from 0.39 to 0.13. It is clearly visible that the predictions based on the AE-GPR approach are more confined than the predictions based on the geometrical approach. Furthermore, the predictor based on the geometrical approach fails clearly in determining the measured nugget diameter on the samples stemming from the subset yielding the smallest nugget diameter. The provided 99 % confidence bounds show a narrow band encapsulating the data.

As a final verification, the algorithm is subjected to several testing points which were excluded from the training. The latter is represented by 10 % of the available data and is visualised by the red crosses. The RMSE for the test points improves from 0.62 to 0.13 for testing data. Both the RMSE and graphical representation of the predicted value indicate a prediction well within the predicted bounds of the algorithm.

5.2 EBR variation

For the final example, a highly relevant industrial case is elected. One of the advantages of many state-of-the-art welding controllers is their ability to monitor themself and intervene during a single weld. This is required to achieve a certain threshold of total current flow and is often realised by extending the weld time or increasing the current. An important reason why this threshold is often not achieved is the variation in the closed loop resistance in the secondary circuit. These variations affect the welding process and potentially the reaction of the controller, causing a variation in the weld nugget diameter. Also, it is well known within RSW industry that the weld nugget diameter is negatively affected by variation in the closed loop resistance in the secondary circuit. The latter is often a consequence of machine disturbances, e.g., maintenance, re-torquing or swapping tools. Thus re-calibration is required when adjustments are made to the welding heads to counter this

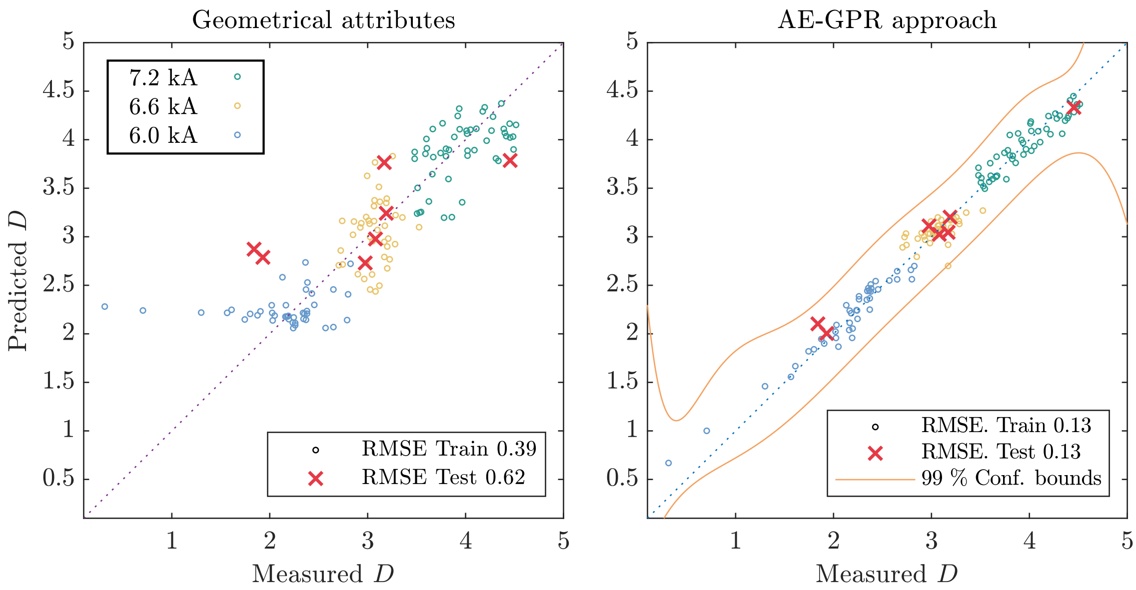

Fig. 10: Predicted vs. actual QI based on the AE-GPR approach *(right)* compared to the Geometrical attributes approach *(left)* for the current variation case

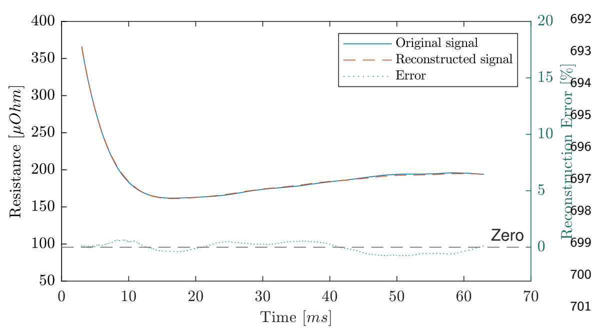

Fig. 11: Measured and reconstructed signal with means of an autoencoder network, with topology $450k - 15 - 450k$ for the current variation case

effect. The goal of this third case is to demonstrate the potential of our approach when the aforementioned effects are present.

The dataset is a combination of three consecutive experimental runs altering the resistance of the secondary circuit by means of custom shims between the welding head and electrode holder, while the compensation of the controller is disabled. The resistance between two fixed points in the mechanical assembly is respectively measured to be 11.9, 108.4 and 371.5 $\mu\Omega$. This is achieved by respectively zero, 3 and 5 shims, as illustrated in figure **??**. This variation is a constant resistance during the process and is part of R_m, as elaborated in section 3.1.

Analysing the measured and predicted data, illustrated in figure 12, clearly reveals several meaningful results. First, the influence of the change from 11.9 to 108.4$\mu\Omega$ has an almost negligible effect on the measured nugget diameter with respectively an average of 4.24 and 4.09 mm and equal standard deviation of 0.18 mm. The experiments where the resistance is increased to 371.5$\mu\Omega$ indicate a noticeably smaller nugget diameter, with an average of 3.18 mm and standard deviation of 0.33 mm. Second, applying the proposed dimension reduction by means of a trained autoencoder reduces the dataset from 140.000 samples to 30, while only losing $< 1\%$ of information, based on mean-squared error estimates.

Third, advancing to the results of the prediction model, there is a clear improvement over the geometrical approach. For the set with 11.9 and 108.4 $\mu\Omega$, the variation in nugget diameter is rather low, yet the AE-GPR is capable of predicting the nugget diameter more accurate compared to the geometrical approach. The nugget diameters for the set with 371.5 $\mu\Omega$ range from 2.7

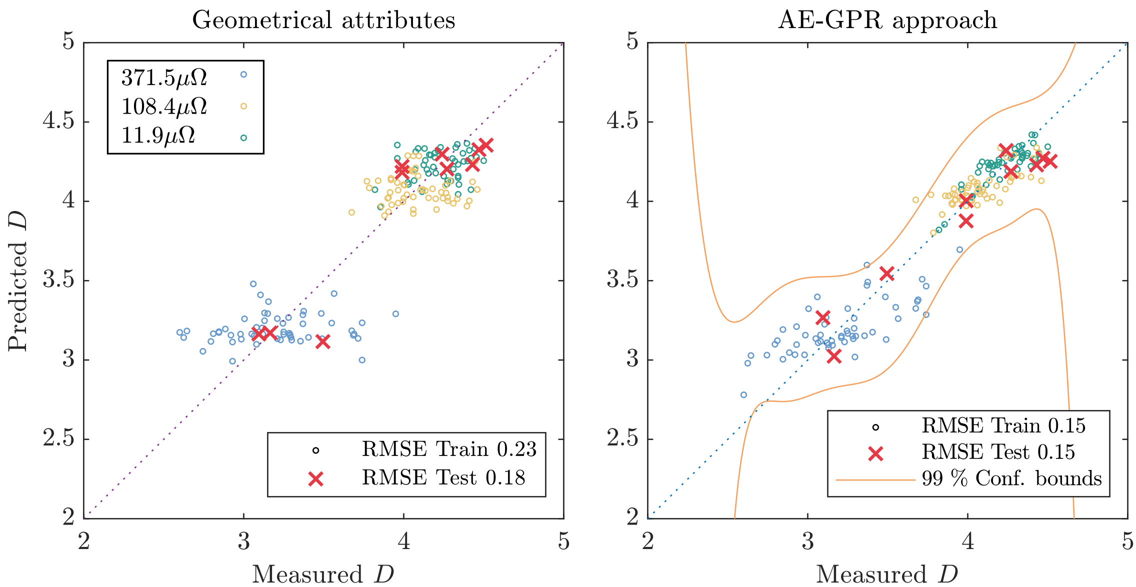

Fig. 12: Predicted vs. actual QI based on the AE-GPR approach *(right)* compared to the Geometrical attributes approach *(left)* for the EBR variation case

to 4.1 mm, while the geometrical approach predicts only between 3.0 and 3.5 mm. The AE-GPR appraoch predicts this set, both for the training and test points. The RMSE of the predictions improves clearly, from 0.23 to 0.15. For this case, the 99% bounds are provided, showing a narrow band encapsulating the data.

The test points are represented by 10 % of the available data. The RMSE for the test points improves from 0.18 to 0.15, resulting in a prediction well within the predicted bounds of the algorithm.

6 Conclusion

This paper presents a novel approach for an effective predictor for the nugget diameter, i.e., the main quality indicator (QI) of a resistance spot welding process, based on on-line measured process data. The most important feature of the developed methodology is the combination of deep learning for dimension reduction, and the consequent machine learning prediction tool. We demonstrate how to use unsupervised deep learning in the form of an autoencoder to discover a low-dimensional transformation in which the parameters characterise a pattern that embodies underlying information on the process, possibly unobservable or not detectable by any other currently existing approach. The underlying information is transformed into a low dimensional space, which is an ideal scene for a Gaussian process regression model linking the input data through the autoencoder to the measured nugget diameter. The model is trained on a limited set of data, leading to a low cost implementation in an industrial setting. The technique is presented in an example case which clearly indicates that it leads to an improved QI prediction compared to the state of the art geometrical attributes approach. The technique is presented on two additional cases, each pinpointing a specific bottleneck within industry related to the RSW process. Both cases are analysed and indicate very promising results, where the new AE-GPR approach has consistently improved results over the geometrical approach. It should be noted that the time to train the proposed model is substantially higher. However, the investment cost is rather low compared to the benefit in an in-line monitoring system. We conclude that the proposed method can be readily applied to an in-line context for quality assessment in industrial RSW applications, as the time for evaluating in an in-line context is suitable, and accuracy is vastly improved over existing techniques.

7 Acknowledgements

This work was funded by KU Leuven, grant #C24E/21/026.
Conflict of interest: The authors declare no conflict of interest.

References

[1] Almeida FA, De Paula TI, Leite RR, et al (2018) A multivariate GR&R approach to variability evaluation of measuring instruments in resistance spot welding process. Journal of Manufacturing Processes 36(November):465–479. https://doi.org/10.1016/j.jmapro.2018.10.030, URL https://doi.org/10.1016/j.jmapro.2018.10.030

[2] Bogaerts L, Faes M, Moens D (2019) A fast inverse approach for the quantification of set-theoretical uncertainty. In: 2019 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, pp 768–775

[3] Brunton SL, Kutz JN (2019) Data-driven science and engineering: Machine learning, dynamical systems, and control. Cambridge University Press

[4] Chen S, Sun T, Jiang X, et al (2016) Online monitoring and evaluation of the weld quality of resistance spot welded titanium alloy. Journal of Manufacturing Processes 23:183–191. https://doi.org/10.1016/j.jmapro.2016.06.003

[5] Cho Y, Rhee S (2002) Primary circuit dynamic resistance monitoring and its application to quality estimation during resistance spot welding. Welding Journal (Miami, Fla) 81(6)

[6] Cho Y, Rhee S (2004) Quality estimation of resistance spot welding by using pattern recognition with neural networks. IEEE Transactions on Instrumentation and Measurement 53(2):330–334

[7] Dejans A, Kurtov O, Van Rymenant P (2021) Acoustic emission as a tool for prediction of nugget diameter in resistance spot welding. Journal of Manufacturing Processes 62(September 2020):7–17. https://doi.org/10.1016/j.jmapro.2020.12.002, URL https://doi.org/10.1016/j.jmapro.2020.12.002

[8] Dickinson DW, Franklin JE, Stanya A (1980) Characterization of Spot Welding Behavior By Dynamic Electrical Parameter Monitoring. Welding Journal (Miami, Fla) 59(6)

[9] El Ouafi A, Belanger R, Guillot M (2012) Dynamic resistance based model for on-line resistance spot welding quality assessment. Materials Science Forum 706-709:2925–2930. https://doi.org/10.4028/www.scientific.net/MSF.706-709.2925

[10] Ghaffari B, Mozurkewich G (2010) Nondestructive evaluation of spot-weld quality. In: Failure mechanisms of advanced welding processes. Elsevier, p 101–136

[11] Gomes GF, Viéville P, Durrenberger L (2017) Dynamic behavior investigation of spot welding machines and its influence on weld current range by modal analysis. Journal of the Brazilian Society of Mechanical Sciences and Engineering 39(3):765–773. https://doi.org/10.1007/s40430-016-0580-0

[12] Hamedi M, Atashparva M (2017) A review of electrical contact resistance modeling in resistance spot welding. Welding in the World 61(2):269–290. https://doi.org/10.1007/s40194-016-0419-4, URL http://dx.doi.org/10.1007/s40194-016-0419-4

[13] Hao M, Osman KA, Boomer DR, et al (1996) Developments in characterization of resistance spot welding of aluminum. Welding Journal (Miami, Fla) 75(1):1–s

[14] Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. science 313(5786):504–507

[15] Ighodaro OLR, Biro E, Zhou YN (2017) Study and Applications of Dynamic Resistance Profiles During Resistance Spot Welding of Coated Hot-Stamping Steels. Metallurgical and Materials Transactions A: Physical Metallurgy and Materials Science 48(2):745–758. https://doi.org/10.1007/s11661-016-3899-3

[16] Krige DG (1951) A statistical approach to some basic mine valuation problems on the witwatersrand. Journal of the Southern African Institute of Mining and Metallurgy 52(6):119–139

[17] Ling SF, Wan LX, Wong YR, et al (2010) Input electrical impedance as quality monitoring signature for characterizing resistance spot welding. Ndt & E International 43(3):200–205

[18] Rasmussen CE, Williams CK (2006) Gaussian Processes for Machine Learning, vol 1. MIT press

[19] Sacks J, Welch WJ, Mitchell TJ, et al (1989) Design and analysis of computer experiments. Statistical science 4(4):409–423

[20] Saxe AM, Bansal Y, Dapello J, et al (2019) On the information bottleneck theory of deep learning. Journal of Statistical Mechanics: Theory and Experiment 2019(12):124,020

[21] Wan X, Wang Y, Zhao D, et al (2017) Weld quality monitoring research in small scale resistance spot welding by dynamic resistance and neural network. Measurement: Journal of the International Measurement Confederation 99:120–127. https://doi.org/10.1016/j.measurement.2016.12.010, URL http://dx.doi.org/10.1016/j.measurement.2016.12.010

[22] Wang SC, Wei PS (2001) Modeling dynamic electrical resistance during resistance spot welding. Journal of Heat Transfer 123(3):576–585. https://doi.org/10.1115/1.1370502

[23] Xia YJ, Su ZW, Li YB, et al (2019) Online quantitative evaluation of expulsion in resistance spot welding. Journal of Manufacturing Processes 46(August):34–43. https://doi.org/10.1016/j.jmapro.2019.08.004, URL https://doi.org/10.1016/j.jmapro.2019.08.004

[24] Xia YJ, Zhou L, Shen Y, et al (2021) Online measurement of weld penetration in robotic resistance spot welding using electrode displacement signals. Measurement: Journal of the International Measurement Confederation 168(September 2020):108,397. https://doi.org/10.1016/j.measurement.2020.108397, URL https://doi.org/10.1016/j.measurement.2020.108397

[25] Xing B, Xiao Y, Qin QH, et al (2018) Quality assessment of resistance spot welding process based on dynamic resistance signal and random forest based. International Journal of Advanced Manufacturing Technology 94(1-4):327–339. https://doi.org/10.1007/s00170-017-0889-6

[26] Zhang H, Hou Y, Yang T, et al (2018) Welding quality evaluation of resistance spot welding using the time-varying inductive reactance signal. Measurement Science and Technology 29(5). https://doi.org/10.1088/1361-6501/aaa830

[27] Zhao D, Wang Y, Lin Z, et al (2013) An effective quality assessment method for small scale resistance spot welding based on process parameters. Ndt & E International 55:36–41

[28] Zhao D, Bezgans Y, Wang Y, et al (2020) Performances of dimension reduction techniques for welding quality prediction based on the dynamic resistance signal. Journal of Manufacturing Processes 58:335–343

[29] Zhao D, Bezgans Y, Wang Y, et al (2021) Research on the correlation between dynamic resistance and quality estimation of resistance spot welding. Measurement 168:108,299

[30] Zhou K, Cai L (2013) Online nugget diameter control system for resistance spot welding. International Journal of Advanced Manufacturing Technology 68(9-12):2571–2588. https://doi.org/10.1007/s00170-013-4886-0

[31] Zhou K, Cai L (2014) Study on effect of electrode force on resistance spot welding process. Journal of Applied Physics 116(8):1–7. https://doi.org/10.1063/1.4893968

[32] Zhou K, Yao P (2019) Overview of recent advances of process analysis and quality control in resistance spot welding. Mechanical Systems and Signal Processing 124:170–198. https://doi.org/10.1016/j.ymssp.2019.01.041, URL https://doi.org/10.1016/j.ymssp.2019.01.041

[33] Zhou L, Zheng W, Li T, et al (2020) A material stack-up combination identification method for resistance spot welding based on dynamic resistance. Journal of Manufacturing Processes 56(March):796–805. https://doi.org/10.1016/j.jmapro.2020.04.051, URL https://doi.org/10.1016/j.jmapro.2020.04.051