



Non-destructive evaluation of the friction stir welding process, generalizing a deep neural defect detection network to identify internal weld defects across different aluminum alloys

P. Rabe¹ · U. Reisgen¹ · A. Schiebahn¹

Received: 29 September 2022 / Accepted: 8 December 2022 / Published online: 9 January 2023
© The Author(s) 2023

Abstract

Friction stir welding (FSW) is a solid-state welding process, which has significantly disrupted welding technology particularly for aluminum alloy applications. Due to its high-quality welds in all aluminum alloys, comparatively low heat input with high energy efficiency and ecological friendliness, FSW is used in a rapidly growing number of applications. Currently, destructive and non-destructive testing methods are attached as a separate process step to verify weld seam quality, detecting imperfections late in production and requiring costly rework or scrapping of the assembly. Various studies have shown the possibility of using deep neural networks (DNN) to evaluate weld quality and detect welding defects based on recorded data. In this study, conducted within the scope of RWTH Aachen's Cluster of Excellence, Internet of Production, recurrent neural networks (RNN), and convolutional neural networks (CNN) were successfully trained to classify FSW force data sets, generated while joining different aluminum alloys over a wide range of welding parameters. For internal weld defects bigger than 0.08 mm, detection accuracies over 95% were achieved using bidirectional long short-term memory (BiLSTM) networks when limited to a single alloy and thickness. The classification accuracy dropped to ~90% when using multiple alloys and sheet thicknesses. The comparison between different network types' classification accuracy as well as their ability to generalize the defect detection across different welding tasks with varying sheet thicknesses, respective welding tools, and different Al alloys is shown. The systems aim at offering a reliable and cost-efficient quality monitoring solution with a wide range of applicability, increasing the acceptance of the friction stir welding process as well as confidence in the resulting weld seam quality.

Keywords Friction stir welding · Quality monitoring · Deep neural network

1 Introduction

Friction stir welding (FSW) is a modern solid-state welding process that produces high-quality welds through intermixing in the plastic state using frictional heat and pressure generated by a rotating tool. The process was patented by the TWI (The Welding Institute) in 1991 [1]. The solid-state nature of the process bypasses challenges associated with fusion welding of aluminum alloys and produces welds with

superior technological properties, making it a sought after joining process in the aerospace and rolling stock industries [2]. Increased efforts in light weight construction as well the growing demand in electric vehicles have increased the number of FSW applications in the automobile industry for the production of battery trays, heat exchangers, and mixed material joints of aluminum and copper for electrical systems [3, 4]. Along with the increases in production volume, the need for reliable and cost-efficient, non-destructive inline quality monitoring, that is easily applicable to the increasing speed of changes in the production environment, is constantly growing across all fields of FSW application [5].

The FSW process is generally implemented on specialized machines capable of highly automated processing using closed loop axial force control to adaptively control the process. While this indicates that sensors to indirectly monitor certain process parameters are already implemented, the

Recommended for publication by Commission V—NDT and Quality Assurance of Welded Products.

✉ P. Rabe
rabe@isf.rwth-aachen.de

¹ Welding and Joining Institute (ISF), RWTH Aachen University, Pontstraße 49, 52062 Aachen, Germany

accuracy and measurement rate vary widely across different manufactures and machine types and generally provide insufficient process feedback for high resolution quality monitoring. Many different approaches using external sensors to establish relationships between weld data and quality control have been published recently, for example, [6–11]. Generally, the quality monitoring is realized through analysis of the dynamic behavior of in plane welding forces or the dynamic variation of axial force or torque. Despite the different approaches on indirect FSW data monitoring, the goal of defect detection is mostly achieved. All of the given examples are limited within their applicability as they only demonstrate feasibility for a certain joining task (one alloy and one sheet thickness), sometimes even limited to one parameter set. This point is addressed within this work in order to demonstrate the generalization and improved applicability of inline capable quality monitoring across different alloys and sheet thicknesses. In order to achieve this goal, deep learning algorithms are used to analyze and categorize the recorded welding data to identify internal weld defects. These internal voids are one of the most difficult defects to detect in FSW and require inspection by phased-array ultrasound or computer tomography [12].

Recent developments of better deep learning algorithms, an increased understanding of machine learning approaches, and easy access to high performance computing increase the potential benefits of recording and analyzing production process data for manufacturers [13]. The used recurrent neural networks (RNN) and convolutional neural networks (CNN) are deep learning algorithms (DL) that include detection and extraction of low-, mid-, and high-level features in the training cycles. This eliminates the need for time consuming, brittle and not scalable hand engineered feature extraction. The automated feature detection and versatile architecture allow for classification of diverse input data [14].

In this work, two types of artificial neural networks (ANN) are used for classification. Both were brought back into focus by recent improvements to architecture, solvers, and activation functions [15, 16] as they perform well computer vision problems (2D data) and time series analysis.

1.1 Typical reasons for FSW weld defects

Compared to most fusion welding processes, the FSW process is regarded as a stable and well controllable process in industrial production environments, allowing for efficient process integration and joining. On the other hand, an implemented, steady-state FSW process is still susceptible to a number of external disturbances that can lead to weld seam defects which are not detectable by the most common quality monitoring procedures, e.g., axial force monitoring and visual inspection coupled with selective destructive testing [17]. These influence factors are commonly divided

into equipment-based factors, welding parameter factors, and workpiece influences. The equipment-based factors are dependent on the welding machine, the used fixture, and welding tools and can be regarded as constant when investigating weld quality [18]. Tool wear is an exception to this and needs to be closely monitored and included in any analysis, not only in production, but also in research [19]. The main welding parameters are often empirically developed and fixed for a given welding task, thereby they do not cause process disturbance by themselves, but rather interact with the workpiece to cause defect-inducing deviations. The most common variations are gap tolerances, thickness variations, surface condition changes, and tool wear [20]. The material volume underneath the welding tool and in consequence for force controlled processes the plunge depth and transport of plasticised material are influenced by gap tolerances, sheet thickness variation, and material strength differences in the production run [20], leading to process instabilities. Tool wear and the surface condition of the workpiece [21] influence the interface condition and material transport through wear and adhesion. Combinations of the mentioned influences cause instability of the process and can lead to a process state in which the employed quality control cannot detect the defects developing in the weld seam.

Two main mechanisms that cause defects in FSW have been identified. A change in plunge depth can be identified readily through machine parameters and related to weld quality. This can be caused by various factors and lead to either insufficient plunge depth that results in incomplete penetration and a decrease in mechanical properties [22], or increased plunge depth. The latter causes close proximity or even direct contact between welding tool and backing plate, causing various defects from adhesions between workpiece and backing plate to tool failures.

The second cause of defect initiation is more difficult to detect as it is based on irregular material flow within the stir zone. Changes in energy input, surface condition, workpiece strength/thickness, interface condition, or tool wear can negatively influence the cyclical process condition and resulting material transport around the tool and thereby disrupt the weld seam formation [23]. This can lead to local or prolonged internal weld seam defects such as voids, cavities, or surface defects [24]. Therefore, with the current state of the art, additional quality assessment is required for FSW production, which increases complexity, manufacturing time, and costs.

1.2 FSW process monitoring and weld seam quality assessment

The FSW process is often characterized by its distinctive and comparatively high process forces. The forces in all three spatial directions are composed of a static and a dynamic

component, which in its steady state presents in a cyclical manner, corresponding to the rotational speed of the welding tool fixed to the spindle [25]. While the periodic nature of the spatial forces is widely agreed upon, the reasons and driving causes are still not fully identified, especially regarding in plane forces and their interdependencies and relation to the welding parameters, tools, and workpiece properties. Despite this uncertainty, many researchers have used various approaches to correlate deviation from uniform force or torque oscillation to weld defects, e.g., [6, 10, 11, 26–29].

The feasibility of empirical force evaluation based approaches was shown in a number of works. These empirical approaches often reduce anomaly detection to simple incremental features or gradual changes along the weld. Jene correlates the mean average lateral force (F_y) to weld seam defects [27], while most other works focus on changes in the dynamic components or combinations of changes in the static and dynamic components. Hattingh et al. [29] monitor forces and torque to determine the field of plasticised material and calculate material flow and resulting flow defects. Luhn [10] graphs the axial force over spindle torque as well as the forces in the welding plane (F_x and F_y) to identify weld seam features. Smooth, uniform graphs indicated defect free welds, while scattered plots indicated weld seam defects. Based on this approach, feed force oscillation (F_x) was identified as an indicator for internal weld defects [11]. It was also shown that a threaded pin greatly increases oscillation in defective welds [11]. The research shown above enables process force-based defect detection but also highlights that empirical monitoring approaches require adaption to each change in the welding condition, thereby limiting applicability.

Besides the welding forces, the frequency spectra can be analyzed to determine voids. Gebhard [18] found that high amounts of low frequency oscillation indicate internal void defects. Frequency domain data was also used for the first ANN defect detection in FSW. Boldsai Khan et al. [30] transformed torque recordings sampled at 51.2 Hz into the frequency domain and used them to train multiple ANNs. A fully connected artificial neural network (FCNN) was able to correctly identify all internal weld defects in their entire testing set, which was strongly biased towards defect free welds. Later, works by the main authors show ANN-based evaluation of wavelet transformations of the in plane forces [31]. The trained FCNN was able to correctly identify internal defects > 0.08 mm with an accuracy of 95%. The welds were labeled through cross section analysis.

The emerging class of convolutional neural networks (CNN) was used by Hartl et al. [8] to classify surface images as well as weld data. Varying welding parameters were used to generate 120 welds, with each weld being split into 17 sections to receive a data set of 2040 classified welds. The highest classification accuracy for internal defect detection

was achieved when using a CNN based on AlexNet [32] architecture and evaluating the lateral force (F_y). Over several training cycles, an accuracy of 79.2% could be achieved.

Many researchers have found different ways to correlate FSW data to weld quality indicators with high accuracy. All of the published works are limited in regard to welding tools and workpiece material and thickness. Furthermore, the examined welding parameters are quite limited and often not applicable to industrial use, due to low productivity. They do however demonstrate the feasibility of using force data to detect internal defects and provide a reliable base to correlate process influences and force deviations in the weld data. On the other hand, they do not provide a reliable base for industrial production as the required training sets for each welding application are large and require extensive effort to correctly produce and prepare.

In this study, the application range of in-line defect detection is increased by using deep learning to examine the temporal sequences of FSW process forces and tool torque and categorize welds with internal defects. A data set is established across a wide range of welding parameters, two different Al alloys and two sheet thickness, with their respective welding tools. The data is classified and labeled through analysis of micro-focus computer tomography (μ -ct) pictures. Deep learning algorithms are then trained to extract features and classify the welds based on force and torque measurements.

2 Experimental setup

2.1 FSW and data acquisition setup

All welding experiments were performed on a FSW machine in portal design with a moving table, built by Precision Technologies Group (PTG) Ltd., type 345C. The machine offers a high structural rigidity with position independent compliance and low vibration excitation response [33]. The workpieces were fixed on a welding table with a form fitting cut out for the backing plate. The backing plate was made of 8 mm mild steel that was ground and artificially oxidized.

The workpieces were $135 \times 330 \text{ mm}^2$ allowing for welds of 300 mm length. Workpieces were cut from sheets made of AA-5754-H22 (1.5 mm and 3 mm thickness) with ultimate tensile strength (UTS) of 240 MPa and AA-7075-T6 (1.5 mm thickness) with UTS of 545 MPa. All welds were performed as one-dimensional blind welds with full thickness penetration in order to circumvent edge and gap influences on the resulting weld seams. The welding parameters used are given in the next chapter. Monolithic welding tools were used to produce the welds. For all of the welding trials, the tool tilt angle was fixed at 1.5° . The pin length was adapted to the sheet thickness to achieve full penetration

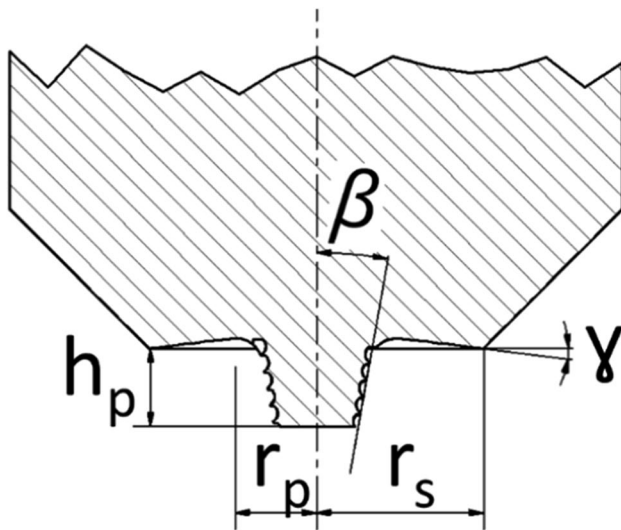


Fig. 1 Tool geometry

Table 1 Tool dimensions

Tool number	Probe radius r_p	Shoulder radius r_s	Conical probe angle β	Probe length h_p	Concave shoulder angle γ
1	2.5 mm	7 mm	10°	1.42 mm	7°
2	2.5 mm	7 mm	10°	2.85 mm	7°

welds. Tool geometry is shown in Fig. 1, and the relevant dimensions are given in Table 1.

The welding data is recorded by a sensor unit integrated into the tool holder, Spike® mobile. The process forces are recorded in three spatial directions F_x , F_y (in welding plane), and F_z , relative to the welding tool. F_x and F_y are recorded as a function of the measured bending moment in the weld

plane relative to the measuring point; the axial force and torque are measures directly through strain gauges within the tool holder. The tool holder, tool, and measured variables for an exemplary milling process are shown in Fig. 2. The data is recorded at a measuring frequency of 2.5 kHz and wirelessly transmitted to a receiver integrated into the machine control system and process control. It is also connected to a computer for processing, recording, and evaluation of the weld data.

2.2 Design of experiments

The welding experiments were designed to analyze the main influences on force response and the resulting deviations in welds containing defects. The main excitation frequency is caused by the tool rotation imprinted on the tool through the spindle speed (RPM), superimposing radial runouts, and discontinuous material transport [34]. The amplitude of the oscillation is mainly due to temperature and volume dependent material resistance to the stirring action. To examine these influences, a wide range of feed rates and tool rotational speeds were used during the experiments. The parameter combinations can be grouped into three distinct sets, determined in order to analyze these influences. Each set consists of five welding parameter combinations with set 2 and set 3 overlapping set 1 in one combination each for a total of 13 different parameter combinations. The parameter combinations are shown in Fig. 3. For set 1, the relationship between feed rate and spindle speed was fixed to two revolutions per millimeter (2 rev/mm), allowing the monitoring of the increase in welding forces due to an increase in feed rate and the associated reduction in welding temperature and thermal softening in front of the welding zone. Sets 2 and 3 are designed around a fixed RPM to keep the oscillation frequency constant and increase the amplitude by increasing the feed rate.



Fig. 2 Spike® instrumented tool holder, data receiver, and processing

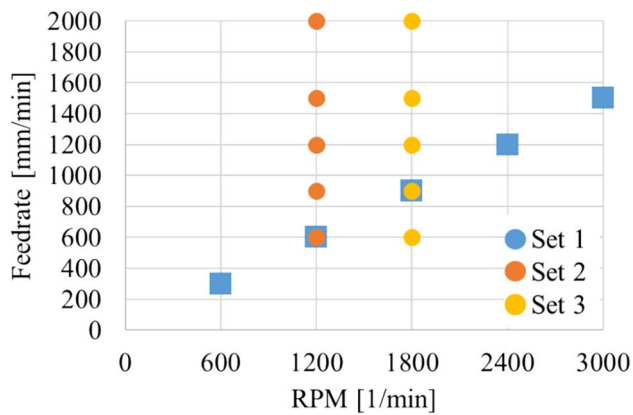


Fig. 3 Welding parameter combinations

The spindle speed for set 2 was chosen as 1200 RPM, which is common in industrial applications and offers a high number of measurements (125) per tool revolution to accurately map the welding forces. The spindle speed of set 3 was set to 1800 RPM, as its first higher harmonic of 60 Hz is equal to the lowest natural frequency of the machine [33] and should, despite pre-tensioning through axial force, slightly increase oscillation, thereby increasing detectable variations in the force signals. The feed rate was varied from 600 to 2000 mm/min for set 2 and set 3. The lower limit was dictated by the minimum rate at which welds could be produced without overheating for both Al alloys at the fixed RPM. The upper limit was set due to load limitations of the used measurement equipment. For set 1, the minimum feed rate could be extended due to the reduced spindle speed and the maximum feed rate was limited by the maximum spindle speed.

Position control mode was chosen as weld control strategy over the more common force control. Position control allows for a reliable plunge depth relative to the machine coordinates and eliminates the influences of varying plunge depth. Furthermore, the influence of the closed loop force controller on the resulting data like lag, force deviation, and machine imprinted oscillation of the z-axis

force is eliminated. This enables the reliable production of defect free as well as defective welds and weld data generation. For each parameter combination, two different plunge depths were determined in pre-trials, resulting in 78 plunge depths for the data set. The first one per parameter combination was set in order to produce defect free welds, and the second depth for each parameter combination, a reduced plunge depth, was chosen to decrease heat generation and forging pressure and thereby produce welds with internal defects, while maintaining defect free weld surfaces.

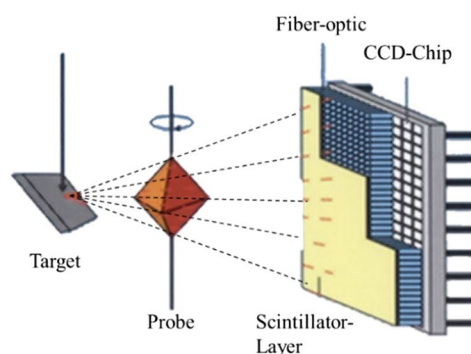
Each parameter combination and plunge depth was repeated three times for each of the alloys and plate thicknesses. With a couple of parameter combinations overloading the measuring device, the number of received weld data sets was reduced from the theoretical 234 welds to 203 welds with recorded sets of weld data.

3 Results and discussion

3.1 Welding results, classification, and data set

The parameter combinations given in Fig. 3 were successfully welded for AA5754 H22 in 1.5 mm and 3 mm sheet thickness and AA7075 T6 in 1.5 mm. The forces of some combinations exceeded the measurement capabilities of the used device thereby reducing the final number of weld data sets to 203 individual welds. The welds were visually inspected for surface defects. No open surface defects were found. The welds were inspected for internal defects using micro-focus computer tomography (μ -ct) [35]. The micro focus allows for magnification of the probe through detector positioning in the opening ray path (Fig. 4) [36]. A duplex wire quality indicator was used to determine the spatial resolution of the generated images [37]. In accordance with the cited ISO standard, the detection threshold for internal volumetric defects (voids and tunnel defects) was determined to be 0.08 mm orthogonal to the plates' surface.

Fig. 4 Micro-focus computer tomography and μ -ct machine Viscom VT 9225



The μ -ct pictures were post-processed to adjust brightness and contrast. The pictures were then analyzed to localize and mark internal defects. Defects and discontinuities at the plunge location were disregarded, as well as steel particle adhesions introduced by the base plate in the same area. Internal weld defects were found in 93 of the 203 welds (45.8%), slightly below the targeted 50%. To validate the findings of the μ -ct inspection, cross sections were taken from selected specimens and analyzed for defects. Figures 5 and 6 show the compounded μ -ct pictures of two sheets with two welds each as well as selected cross sections of the welds, including the extraction plane and viewing direction of the cross sections. The cross section analysis is in good accordance with the μ -ct analysis and can be regarded as reliable for the identified detection threshold for both sheet thicknesses. Small defects of <0.2 mm width and between 0.17 and 0.06 mm height are clearly identifiable in Fig. 6. The μ -ct analysis is therefore used to categorize all welds.

To increase the training set size and offer the possibility of inline quality control, the welds are split into shorter

sections. A new section starts each second with a length of three seconds (3 s) thereby overlapping the previous and subsequent sections. To receive the same number of sections from each weld, the maximum welding speed of 2000 mm/min and weld length were set as the reference, allowing for five sections from each weld. The sections of welds with lower feed rates were taken from the end of the weld sparing the tool exit location. A graph of the sections along with a picture of a weld made at the maximum feed rate of 2000 mm/min is shown in Fig. 7.

The welds are categorized into welds with internal defects (NOK) and welds without any internal defects (OK). The categorization is made over the entire weld length as well as for the shorter sections. This enables the training to be performed on both full length weld data, as well as the shortened overlapping segments. The results of the categorization with the determined threshold of 0.08 mm defect size are shown in Table 2 for the entire sets as well as each individual welding task. The categorization resulted in eight data sets of different sizes and complexities to be used for training, comparison, algorithm generalization, and validation.

Fig. 5 μ -ct of a specimen 1.5 mm AA5754 with different welding parameters including cross sections for validation

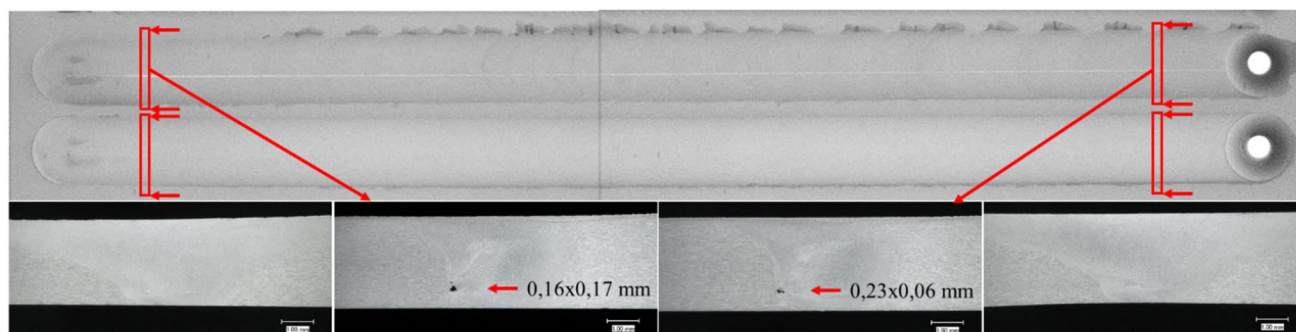


Fig. 6 μ -ct of a specimen 3 mm AA5754 with different welding parameters including cross sections for validation

Fig. 7 Weld at 2000 mm/min and data segments for analysis

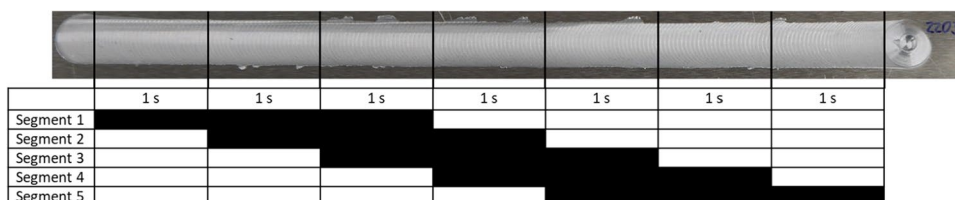
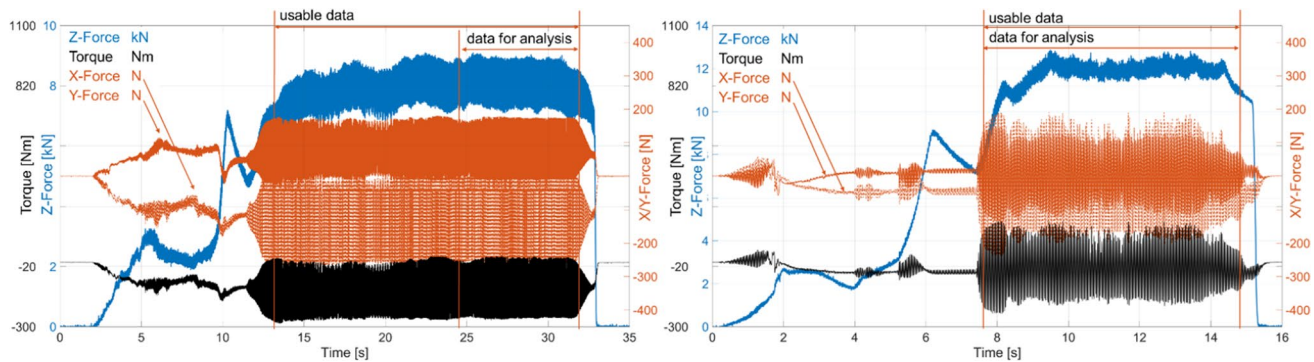


Table 2 Classification of welds and weld segments

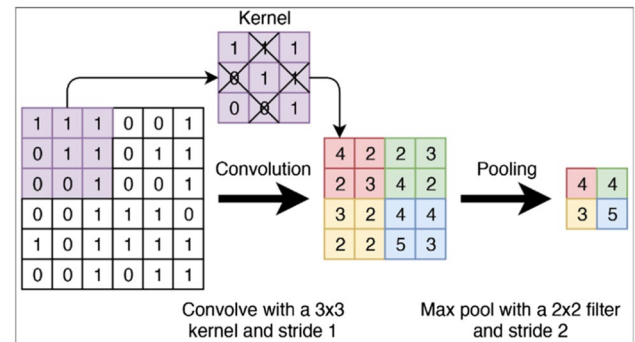
Welds in each category	Full set		AA5754 1.5 mm		AA7075 1.5 mm		AA5754 3 mm	
	Full	Segments	Full	Segments	Full	Segments	Full	Segments
Defect (NOK)	93 (45.8%)	451 (44.4%)	32 (40.0%)	154 (38.5%)	40 (48.8%)	196 (47.8%)	21 (51.2%)	101 (49.3%)
No defect (OK)	110 (54.2%)	564 (55.6%)	48 (60.0%)	246 (61.5%)	42 (51.2%)	214 (52.2%)	20 (48.8%)	104 (50.7%)
Sum	203	1015	80	400	82	410	41	205

**Fig. 8** Recorded data of (left) defect free weld at 600 mm/min and (right) weld with internal defects at 2000 mm/min

3.2 Modeling, training, and testing of the neural networks

The welding data was used to train deep neural networks to detect characteristic features and classify the welding force response data to identify internal weld seam defects. DL networks were chosen for their ability to detect high level features, exceeding the manually identified and implemented gradients, thresholds, or frequency-based empirical features used for quality control. Deep learning methods replace the manual features through multiple layers that weight the inputs (linear combination of selected inputs), sum the weighted inputs, and apply non-linear activation functions to generate an output [14]. For this work, the classification ability of two types of DL networks was examined. The differences in architecture change the way the data is learned and features are generated.

The recorded welding force and torque data were used to train the networks using supervised learning with the categorization described previously. The weld data can also be visualized to validate its quality and determined simple features. As an example of the recorded data, which the ML is based upon, Fig. 8 is shown. The data for axial (Z-force), in plane forces (feed force x and lateral force y), and torque are identifiable on their respectively scaled axis with their static and dynamic components. As described, an equal length of weld data towards the end of the weld is used in training, shown as “data for analysis.”

**Fig. 9** Graphic representation of convolutional filter and pooling layer [40]

The first investigated network type is convolutional neural networks (CNN). The features within the 2D data are detected by banks of convolutional filters of various sizes that slide along the input features. The shared-weight architecture of the filters makes them shift and space invariant providing translational equivariance known as feature maps [38]. This helps to recognize and categorizes patterns across the input data. The network architecture directly influences the complexity of detectable features and patterns a CNN can identify. The detectable feature complexity is proportional to the number of filter layers and their sizes and numbers, as each convolutional layer allows for detection of more complex features. Figure 9 shows a representation of one

very simple filter (size 3×3) with a stride of 1 and another typical layer type, a pooling layer [39, 40].

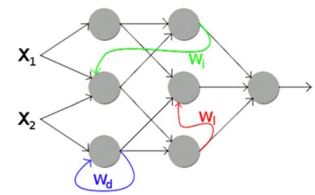
For the CNN application, the recorded data (F_x , F_y , F_z , and torque) of each segment was reshaped to mirror a one-dimensional pixel string, where each recorded data point value represents the grayscale value of a single pixel. This reduces computational time as it combines four separate input streams into one of four times the length, resulting in a data series of 70,000 full weld length or 30,000 for segments (7 s/3 s at 2.5 kHz * 4 data streams) by 1 value, and the binary classification was used as the output. Based on previous works [26], an architecture based on AlexNet [41] was chosen to maximize the feature complexity and detection accuracy. The original architecture was adapted by replacing the rectified linear unit (ReLU) activation functions through exponential linear units (ELUs) to enable negative values and thereby push mean unit activations closer to zero. This enables faster learning rates and enables better generalization in networks containing more than five layers [16]. ELUs are also better at batch normalization than ReLUs, which is important for big data sets and graphics processing unit (GPU) acceleration, as video memory is limited and smaller batch sizes need to be employed when utilizing consumer hardware. A block diagram of the resulting layer architecture is shown in Fig. 10. Filter numbers, sizes, and strides of the network were optimized throughout the training cycles to maximize test-set classification accuracy.

The second network type used was recurrent neural networks (RNNs) that were designed to learn long- and short-term dependencies of time-series data. To do this, the network forms a graph between its nodes along the temporal sequence, exhibiting weighted dynamic temporal behavior [42].

Figure 11 shows a number of possible sequences. Further information on architecture and operation principles can be found in [42, 43].

The RNNs were trained with four input neurons (F_x , F_y , F_z , and torque) and the binary classification as the output. The RNNs contained a bi-directional long short-term memory layer (BiLSTM), which equates to two long short-term memory layers, one learning the dataset forward, the other

Fig. 11 RNN with two inputs one output and direct (blue), indirect (green), and sideways (red) feedback



one backwards. The number of hidden units in the BiLSTM layer was varied to find the optimal relationship between classification accuracy and overfitting prevention. The hidden layers were followed by one fully connected layer with two outputs, to match the number of output classes, one softmax layer and finally the classification layer.

The used dataset was randomly divided for each training setup (each iteration of CNN and RNN) into 80% training data, 10% validation data, and 10% test data. The training data is used to learn the features and calculate weights and biases for the network. These weights and biases are checked and reset during training using the validation data at specified intervals. The fully trained network is then applied to the test set, which data was not used during training to test the classification accuracy of the generated network. The resulting classification accuracies depend on the random distribution of data among the separate sets as well as the weight initialisation. Therefore, each setup is repeated three times ($n=3$) to validate the results and prevent bias due to unbalanced datasets (OK/NOK) and outliers due to initialisation bias. Furthermore, the order of the data sets was shuffled before each training cycle. For the training of RNNs and CNNs, the ADAM optimizer [44] was used at different, constant learning rates.

3.3 Comparison of neural network architecture classification accuracy, generalization ability, and computational requirements

To analyze the gathered data and categorize the welds according to high-level features and short time dependency evaluation, CNNs based on the AlexNet architecture were

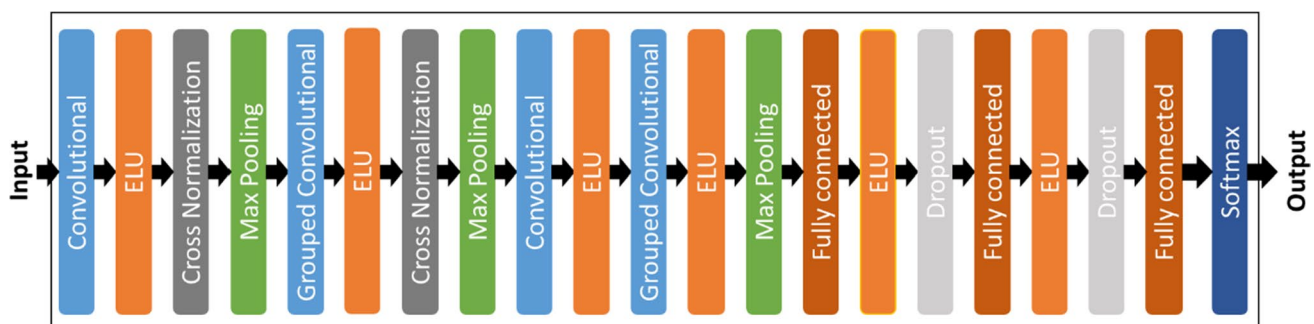


Fig. 10 Architecture of adapted AlexNet CNN network

used. The intricate architecture of layers of convolutional filters and grouped convolutions allows the detection of multiple features in parallel (grouped convolutions) and complex features (multiple layers). The filter counts and sizes were reduced for computational ease as the number of recognizable features and resulting classifications in this work is vastly lower than in the original network design purpose. All the data (F_x , F_y , F_z , and torque) was reshaped into a one-dimensional string of values by stacking the individual inputs, corresponding in length to the duration and measurement frequency of the four stacked inputs (i.e., $70,000 \times 1$ and $30,000 \times 1$ data points). For each data set or subset, the input layer was adjusted accordingly and various setups of filter size, number, and stride were tested. Special focus was given to the relationship between the initial filter size, its stride, and the relationship to the number of data points per revolution of the two main excitation frequencies (measurement frequency/spindle speed). The achieved test-set classification accuracies for the best filter setups are shown in Table 3 for network training with the full test set for both full length and shortened, overlapping segment data (Fig. 7), as well as individual training with segments for the different welding tasks, separated by Al alloy and sheet thickness. The resulting categorization accuracy of the CNN is satisfactory for the full-set maximum length data, as well as the individual data sets for both thicknesses of AA5754, reaching over 81% (49 of 60) on the full data set and ~90% for AA5754 subsets. Contrasting, these results are the significantly worse results for the segments of the entire set; despite the increased number of observations and overlapping data, only 75.49% of test data was classified correctly and only 74.80% of the AA7075 test-set data was classified correctly. No overfitting occurred during the training due to the chosen architecture and the classification accuracy of the test-set data matched the training accuracy well. During the analysis of the results, no clear indications could be made towards false positives or false negatives as their prevalence shifted by each iteration and test-set allocation. Regarding the joining task (alloy and thickness), it was found that the results of the full-set training reflect the results of

the individual training, thereby performing worse for the data corresponding to welds of AA7075 and better for data from AA5754 welds. Overall, it can be seen that the CNN can generalize across sheet thickness and alloy but does lose classification accuracy, especially when categorizing shortened segments. Individual networks with adjusted filter configurations outperformed the generalized network at the cost of computational time for training.

The second network type was RNNs with a bi-directional long short-term (BiLSTM) teachable layer. The BiLSTM RNN was used for its main advantage of learning long-term dependencies from time-series data. The three spatial forces and tool torque (F_x , F_y , F_z , and torque) were used as parallel input sequences ($4 \times 17,500$ for full length and 4×7500 data points for segments). The number of hidden units in the layer was adapted for each training subset to optimize classification accuracy without overfitting to the training data. Separate setups with multiple (2–3) BiLSTM layers and interjected dropout layers to prevent overfitting were investigated but could not reliably improve accuracy while significantly increasing computational time. Unlike previous works based on different data sets [26], the BiLSTM RNNs perform better than the CNNs for the full data set as well as across all subsets, presumably benefitting from the increase in data set size (144 to 1015 segments). The benefit of increased data can be seen when comparing segment accuracy of the full set, to the full length data, with an increase in accuracy of 1.44% from 88.56% to 90.00% (see Table 3). The RNN shows a high level of generalization ability, only slightly outperformed by the CNN in cross alloy and thickness feature detection, while performing at a significantly higher accuracy level. The average classification accuracy of the individual subsets increases by 2.81% compared to 2.32% of the CNN when weighted by the number of test-set samples. Due to the varying length of the used data and different numbers of observations in each data subset, the number of hidden units in the teachable layer varied between 200 (205 segments of 3 mm AA5754) and 525 for the full set, full length training. Figure 12 shows an example of the 3 mm AA5754 training. A quick initial convergence can be seen, with asymptotical training accuracy convergence over

Table 3 Classification accuracy of the respective test data sets for the DNNs for the full training set and subsets

Classification accuracy		Data set size	Test-set size 10% data * $n = 3$	Acc. CNN $n = 3$	Acc. BiLSTM $n = 3$
Full set	Max length (7 s)	203	60	81.67%	90.00%
	Segments (3 s)	1015	306	75.49%	88.56%
5754 1.5 mm	Segments (3 s)	400	120	91.67%	96.67%
7075 1.5 mm	Segments (3 s)	410	123	74.80%	88.62%
5754 3.0 mm	Segments (3 s)	205	63	87.30%	93.65%
Average classification accuracy for individual training				84.59%	92.98%
Weighted (number samples) average classification accuracy				83.99%	92.81%

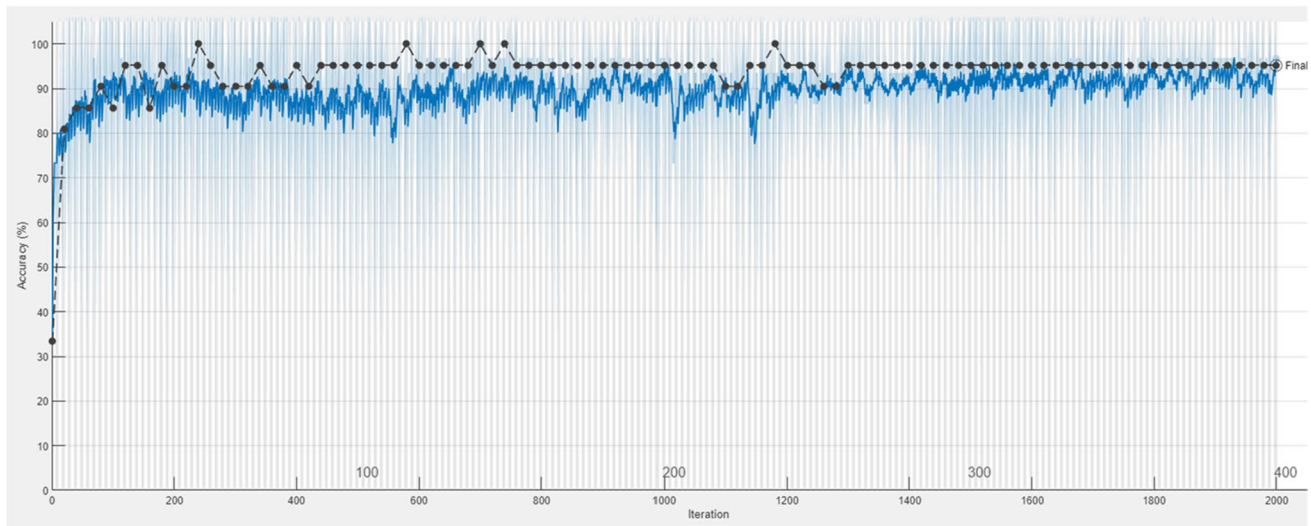


Fig. 12 RNN BiLSTM training iterations of AA5754 3 mm data

the full training cycle. Due to limited GPU memory, the training batch size had to be small (here, 33 weld segments) and therefore shows periodically varying accuracy.

Analog to the CNNs, no clear indication towards false positives or false negatives could be found for the test sets over the iterations. In individual training, the AA5754 1.5 mm data delivered the best network for classification, exceeding 96% over three iterations. The data from the AA7075 welds again proved the most difficult for classification, resulting in a test-set accuracy of 88.62%.

When training DL networks based on AlexNet CNN architecture and BiLSTM RNNs, the RNNs outperform the CNNs significantly in classification accuracy for the investigated filter and layer setups. Generalizing over the data set including different alloys and sheet thicknesses, the RNNs correctly classify 8.33% more of the test data, increasing to 13.07% when classifying 3 s long segments of the data set. This trend can also be seen when each alloy and thickness is evaluated individually, resulting in a classification difference of 8.82%. Along with previous results and studies by other authors, it is to be expected that the CNN could be even further improved to fit the data set and deliver higher accuracies, rivaling the RNN classification performance.

The achieved higher classification accuracy of the RNNs went along with reduced computational requirements. The complex architecture of the CNN leads to expansive data sizes during training. Depending on batch size and filter arrangement, the net and training data exceeded 80 gigabyte (GB) during training. The filter setup strongly influenced the training duration and needed iterations for convergence. On a high performance compute-cluster, the network training took between 12 and 36 h utilizing two Intel Xeon Gold 6258R (28 cores each at 2.7 GHz base clock) at 50% load

and enough volatile memory (RAM) to store all required data. The resulting classification nets are about 2.7 GB in size and once loaded allow for weld data classification in < 0.02 s on a mobile computer (Intel i5 6200U dual core, 2.3 GHz, 15 W TDP, RAM > network size). The reduced size of the RNN during training allowed for GPU acceleration of the matrix computations. A consumer GPU (Nvidia RTX3090 24 GB) was used, reducing training time to 4–12 h depending on input data size and the number of hidden units in the BiLSTM layer. The resulting classification networks are ~ 5 MB in size and allow for weld data classification in < 0.01 s on a mobile computer.

4 Conclusion

For this work, 203 welds with different welding parameters, Al alloys, and sheet thicknesses were produced. The welds were classified according to μ -ct pictures into defect free welds and welds containing inner defects > 0.08 mm. The recorded torque values and welding forces in three spatial directions relative to the welding tool were used to train different architectures of DL networks, BiLSTM RNNs, and CNNs. The networks were investigated regarding their classification accuracy and generalization ability for longer and shorter weld seam segments.

- It is possible for various DL architectures to identify classifying features in weld data without the need for data pre-processing and explicit feature generation.
- The force feedback of physical phenomena and process disturbances during FSW can be used to classify welds

with internal defects without explicit feature identification.

- The investigated DL architectures were able to classify welds from the data set as well as subsets based on force and torque recordings.
- The investigated DL architectures were able to generalize the feature recognition and classification across data set from multiple Al alloys and sheet thicknesses.
- For both RNN and CNN, the classification performance dropped measurably when generalizing across different alloys and sheet thicknesses.
- The classification of segmented welds provides a base for online quality monitoring based on DL of FSW force data.
- RNNs outperformed CNNs in classifying test data of the entire set as well as all investigated subsets. The generalization ability of CNNs exceeded RNNs in testing.
- RNNs achieved classification accuracies of 90% on the entire test set and up to 96% on alloy and thickness specific subsets.

The work proved the viability of using DL networks to identify internal weld seam defects in FSW as a means for inline quality control. For the established data set and μ -ct analysis based classification, it can be summarized that RNNs are better suited for training and detection of internal weld seam defects than CNNs.

Furthermore, the categorization accuracy reduction due to the generalization for both investigated network types leads to the conclusion that supplemental information integrated within the data pre-processing or training stages will improve classification accuracy. The supplemental data should relate to the physical properties of the weld material and its behavior under processing conditions. This will enable the data normalization of the training cycle to better generalize the recorded data to make features signifying weld defects more comparable for the network training.

In a next step, the option of reinforcement learning to categorize further alloys and sheet thicknesses based on the learned features of the trained ANNs will be evaluated.

Funding Open Access funding enabled and organized by Projekt DEAL. The work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy–EXC 2023 Internet of Production–390621612.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source,

provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Thomas W.M.: Improvements relating to friction welding. European Patent Specifications 0615 48 B1
2. Lohwasser D (Hrsg.) (2010) Friction stir welding. From basics to applications. Woodhead Publishing in materials. Boca Raton, Fla., Oxford: CRC Press; WP Woodhead Publ
3. Richter B (2017) Robot-based friction stir welding for E-mobility and general applications. *Biuletyn Instytutu Spawalnictwa* 2017(5):103–110
4. Sharma N, Khan ZAU, Siddiquee AN (2017) Friction stir welding of aluminum to copper—an overview. *Trans Nonferrous Metals Soc China* 27(10):2113–2136
5. Taheri H, Kilpatrick M, Norvalls M, Harper WJ, Koester LW, Bigelow TU, Bond LJ (2019) Investigation of nondestructive testing methods for friction stir welding. *Metals* 9(6):624
6. Boldsai Khan E, Logar AMU, Corwin EM (2010) Real-time quality monitoring in friction stir welding. The use of feedback forces for nondestructive evaluation of friction stir welding. Saarbrücken: Lambert Academic Publishing
7. Das B, Pal SU, Bag S (2016) A combined wavelet packet and Hilbert-Huang transform for defect detection and modelling of weld strength in friction stir welding process. *J Manuf Process* 22:260–268
8. Hartl R, Bachmann A, Habedank JB, Semm TU, Zaeh MF (2021) Process monitoring in friction stir welding using convolutional neural networks. *Metals* 4:535
9. Mishra D, Roy RB, Dutta S, Pal SKU, Chakravarty D (2018) A review on sensor based monitoring and control of friction stir welding process and a roadmap to Industry 4.0. *J Manuf Process* 36:373–397
10. Luhn T (2013) Prozessdiagnose und prozessüberwachung beim rührreißschweißen. Zugl.: Ilmenau, Techn. Univ., Diss., 2012. Berlin: Pro Business 2013
11. Rabe P, Schiebahn AU, Reisgen U (2021) Force feedback-based quality monitoring of the friction stir welding process utilizing an analytic algorithm. *Welding in the World* 65(5):845–854
12. Friction stir welding (2010) Chapter 9. Elsevier 2010
13. Wuest T, Weimer D, Irgens CU, Thoben K-D (2016) Machine learning in manufacturing: advantages, challenges, and applications. *Prod Manuf Res* 4(1):23–45
14. Alexander Amini (2021) Introduction to deep learning MIT Course
15. Lindemann B, Müller T, Vietz H, Jazdi NU, Weyrich M (2021) A survey on long short-term memory networks for time series prediction. *Procedia CIRP* 99:650–655
16. Fast and accurate deep network learning by exponential linear units (ELUs), Clevert, D.-A., Unterthiner, T. u. Hochreiter, S., 2015
17. Mishra RS, De PSU, Kumar N (2014) Friction stir welding and processing. Science and engineering. Cham, Heidelberg: Springer
18. Gebhard P (2011) Dynamisches Verhalten von Werkzeugmaschinen bei Anwendung für das rührreißschweißen. Zugl.:

- München, Techn. Univ., Diss., 2010. Forschungsberichte / IWB, Bd. 253. München: Utz 2011
19. Hattingh DG, Blignault C, Niekerk TIU, James MN (2008) Characterization of the influences of FSW tool geometry on welding forces and weld tensile strength using an instrumented tool. *J Mater Process Technol* 203(1–3):46–57
 20. Cole EG, Fehrenbacher A, Shultz EF, Smith CB, Ferrier NJ, Zinn MRU, Pfefferkorn FE (2012) Stability of the friction stir welding process in presence of workpiece mating variations. *Int J Adv Manuf Technol* 63(5–8):583–593
 21. Więckowski W, Burek R, Lacki PU, Login W (2018) Analysis of wear of tools made of 1.2344 steel and MP159 alloy in the process of friction stir welding (FSW) of 7075 T6 aluminium alloy sheet metal. *Eksploatacja i Niezawodność - Maintenance and Reliability* 21(1):54–59
 22. Muhayat N, Zubaydi A, Sulistijono U, Yuliadi MZ (2014) Effect of tool tilt angle and tool plunge depth on mechanical properties of friction stir welded AA 5083 joints. *Adv Appl Mechanics Mater* 493:709–714
 23. Zettler R, Lomolino S, dos Santos JF, Donath T, Beckmann F, Lippman TU, Lohwasser D (2005) Effect of tool geometry and process parameters on material flow in FSW of an AA 2024–T351 alloy. *Welding World* 49(3–4):41–46
 24. Deutsches Institut für Normung: DIN EN ISO 25239–5, Friction stir welding - aluminium. Part 5, Quality and inspection requirements (ISO/DIS 25239–5:2019). = Rührreibschweißen - Aluminium. Teil 5, Qualitäts- und Prüfungsanforderungen (ISO/DIS 25239–5:2019). Deutsche Norm. Berlin: Beuth Verlag GmbH 2019
 25. Franke D, Rudraraju S, Zinn MU, Pfefferkorn FE (2020) Understanding process force transients with application towards defect detection during friction stir welding of aluminum alloys. *J Manuf Process* 54:251–261
 26. Rabe P, Schiebahn AU, Reising U (2022) Deep learning approaches for force feedback based void defect detection in friction stir welding. *J Adv Joining Process* 5:100087
 27. Jene T (2008) Entwicklung eines verfahrens zur prozessintegrierten prüfung von rührreibschweißverbindungen des leichtbaus sowie charakterisierung des ermüdungsverhaltens der fügen. Zugl.: Kaiserslautern, Techn. Univ., Diss., 2008. Werkstoffkundliche Berichte, Bd. 21. Kaiserslautern: Techn. Univ. Lehrstuhl für Werkstoffkunde
 28. Roberts J (2016) Weld quality classification from sensory signatures in friction-stir-welding (FSW) using discrete wavelet transform and advanced metaheuristic techniques. LSU Master's Theses
 29. Hattingh DG, van Niekerk TI, Blignault C, Kruger GU, James MN (2004) Analysis of the FSW Force footprint and its relationship with process parameters to optimise weld performance and tool design. *Welding World* 48(1–2):50–58
 30. Boldsai Khan et al (2006) Proceedings of 6th international friction stir welding symposium. 10 - 13 October 2006, Saint Sauveur, Canada. Cambridge: TWI Ltd 2006
 31. Boldsai Khan E, Corwin EM, Logar AMU, Arbogast WJ (2011) The use of neural network and discrete Fourier transform for real-time evaluation of friction stir welding. *Appl Soft Comput* 11(8):4839–4846
 32. Wei, J (2019) AlexNet: the architecture that challenged CNNs. *Towards Data Science*
 33. Rabe P, Motschke T, Schiebahn AU, Reising U (2020) Methode zur umsetzung von rührreibschweißprozessen auf konventionellen fräsmaschinen mittels eines empirischen ansatzes. *Schweißen und Schneidejn* 72:1–2
 34. Ambrosio D, Wagner V, Dessein G, Paris J-Y, Jlaiel KU, Cahuc O (2021) Plastic behavior-dependent weldability of heat-treatable aluminum alloys in friction stir welding. *Int J Adv Manuf Technol* 117(1–2):635–652
 35. Kerckhofs G, Schrooten J, van Cleynenbreugel T, Lomov SVU, Wevers M (2008) Validation of x-ray microfocus computed tomography as an imaging tool for porous structures. *Rev Sci Instruments* 79(1):13711
 36. Viscom: X-ray tubes, 2021. <https://www.viscom.com/en/products/x-ray-tubes/>, last downloaded: 16.11.2022
 37. ISO 19232–5. Non-destructive testing — image quality of radiographs. Determination of the image unsharpness and basic spatial resolution value using duplex wire-type image quality indicators
 38. Zhang W, Itoh K, Tanida JU, Ichioka Y (1990) Parallel distributed processing model with local space-invariant interconnections and its optical architecture. *Appl Optics* 29(32):4790–4797
 39. Kuo C-CJ (2016) Understanding convolutional neural networks with a mathematical model. *J Vis Commun Image Represent* 41:406–413
 40. Miguel Fernandez Zafra (2020) Understanding convolutions and pooling in neural networks: a simple explanation. A visual explanation on the concepts that make convolutional neural networks work and the intuition behind them. *Towards Data Science*
 41. Krizhevsky A, Sutskever IU, Hinton GE (2017) ImageNet classification with deep convolutional neural networks. *Commun ACM* 60(6):84–90
 42. Tealab A (2018) Time series forecasting using artificial neural networks methodologies: a systematic review. *Future Comput Inform J* 3(2):334–340
 43. Grossberg S (2013) Recurrent neural networks. *Scholarpedia* 8(2):1888
 44. Adam (2014) A method for stochastic optimization, Kingma, D. P. u. Ba, J

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