# Federated Learning for Weld Quality Prediction

Shubhranil Chakraborty[1], Sayani Guha[1], Debasish Mishra[2], Surjya K Pal[3]*

[1]Department of Electronics and Communication Engineering, Jalpaiguri Government Engineering College, India
[2]Data Science and Information Systems, Krea University, India
[3]*Department of Mechanical Engineering, Indian Institute of Technology Kharagpur, India

*Corresponding author:

Surjya K Pal
Email: skpal@mech.iitkgp.ac.in
ORCID ID: https://orcid.org/0000-0003-2182-6349

This study introduces a monitoring system that can accurately predict the quality of friction stir welds. The study involved four different sets of welding experiments using varying materials and tool configurations. The goal was to create a universal monitoring system capable of predicting weld quality across these diverse experimental sets. Real-time welding data was collected using instruments such as load cells and a power sensor. Signal processing methods were then used to analyze this data and extract essential information about welding quality. Subsequently, Federated Learning (FL) was used to develop the universal monitoring system. This method involves collective learning and model training, leading to a global model trained on decentralized data from the different experimental sets. The approach proved to be effective in forecasting weld quality, achieving a mean absolute error (MAE) of 10.44 for all welding experiments. Additionally, it offered benefits such as reduced latency and enhanced user data protection, while maintaining the accuracy of the global model. An artificial neural network (ANN) model was also developed for comparison with FL, achieving a MAE of 13.85 for the welding experiments. Overall, this study demonstrates the effectiveness of training global and more reliable models, with multiple devices sharing their knowledgebases to train the global model effectively.

**Keywords**: Federated learning; Deep learning; Manufacturing; Friction stir welding; Neural network; Industry 4.0; Quality

## Introduction

In recent years, there has been a significant development in the use of machine learning (ML) systems to identify irregularities in manufacturing activities [1]. These systems, powered by ML algorithms, offer several advantages, including the capability to handle complex and high-dimensional data, solve non-deterministic polynomial problems, and optimize discrete activities [2], [3]. In addition, neural networks (NN) have emerged as effective tools for modeling manufacturing activities, as they can automatically learn features and recognize patterns in sensory data related to manufacturing [4], [5]. Unlike traditional ML approaches that rely on human knowledge for feature extraction, NN-based solutions are designed end-to-end, resulting in higher accuracy.

However, it is important to note that NN-based solutions may require more time and computation resources. Deep learning (DL) solutions are also crucial, as they aid in the detection of minor errors, leading to cost reductions that directly impact industry capital [6]. Researchers have successfully utilized NNs to model, optimize, and predict various parameters in manufacturing processes.

For example, a backpropagation neural network (BPNN) was employed to predict the flank wear in turning operations [7]. A study on surface roughness prediction utilized an artificial neural network (ANN) model [8]. Furthermore, the analysis of drill flank wear employed a fuzzy logic backpropagation algorithm [9]. Additionally, a multi-sensory-based monitoring approach was used to model characteristics of welding processes like weld deposition and plate distortion in pulsed metal inert gas welding [10]. The faults of a reactor tank were classified using a convolutional neural network (CNN) model developed with features extracted using the Andrews function [5]. For further details on the applications of various ML and DL algorithms in different manufacturing processes, readers can refer to this specific article [11].

In the realm of manufacturing, welding is a crucial process in the creation of structures, machines, and assemblies. Typically occurring at the end of the manufacturing process, ensuring the integrity of the weld is of utmost importance. While industries commonly employ non-destructive methods to guarantee the quality of welded products, these methods often pose the drawbacks of being both time-consuming and costly. To mitigate these issues, alternative indirect methods are being opted that involve sensors to collect welding data. These methods harness computer applications such as artificial intelligence (AI), ML, and DL algorithms to analyze sensory data and automate the welding process [12].

The current study presents an application of a DL model specifically developed for the friction stir welding (FSW) process. This type of welding is promising for diverse industries due to its ability to weld workpieces in both similar and dissimilar configurations. The process involves the plastic deformation of the workpiece using a high-speed rotating tool and stirring of the plasticized material as a result of the relative motion between the tool and the machine bed, ultimately leading to welding. To realize the industrial applicability of this process, it was imperative to develop models that could forecast the quality of the welded products.

Several studies have researched modeling the intricacies of the FSW process using a wide range of ML algorithms and NNs. In one study, a Best First Tree (BFT) classifier was developed to predict the condition of the tool used for FSW by analyzing statistical information derived from the vibration signals [13]. Another study focused on creating an ANN model to predict weld strength using features extracted from the current signals [14]. Another study found that the NN surpassed response surface methodology (RSM) in predicting weld strength, demonstrating the superior performance of the NN model [15]. Furthermore, another study explored various algorithms including ANN, RSM, support vector machines (SVM), and adaptive neuro-fuzzy inference systems (ANFIS) to optimize FSW and friction stir spot welding (FSSW) processes [16]. This study revealed that the ANN model was proved to be the most effective due to its ability to recognize patterns in nonlinear data. Additionally, another study reported that modeling using ANN outperformed Gaussian process regression (GPR), SVM, and linear regression (LR) [17]. Furthermore, in another study, two ANN models were developed for FSW, where the first ANN predicted weld strength, while the second ANN predicted crucial welding parameters to control the process [18], [19]. For an in-depth review of sensor and ML-based modeling of FSW, readers can refer to this specific study [20].

The literature contains numerous studies that have explored a variety of sensors used for collecting data on FSW. These studies have also employed a range of ML and DL algorithms to model the welding characteristics. However, a major limitation of these studies is that the ML or DL systems have been developed and validated on data collected from a single source, specifically the source of data collected during welding experiments. This limitation hinders their applicability to new devices that may differ from the training data, which is particularly problematic in the context of welding activity due to variations in materials and configurations.

To overcome these challenges, this study has developed a universal model using the Federated learning (FL) approach. The study involved conducting four different sets of welding experiments, each with varying material and tool configurations, to create a universal monitoring system capable of predicting weld quality despite differences between the experimental sets. Real-time welding data was collected using instruments such as load cells and a power sensor. The FL approach utilizes data from these diverse sources to train the universal model, ensuring its adaptability to changing welding configurations and eliminating the need to store data from individual devices in the cloud. Additionally, the study has also developed an ANN model, providing a valuable point of comparison with the FL approach. The main contributions of this study are as follows:

1. Application of the Federated learning (FL) approach for the accurate prediction of weld strength.
2. Development of a universal monitoring system using the FL approach to predict weld strength across devices.
3. Comparative study with an ANN model in predicting weld strength.

# Experiments

## Welding Experiments and Data Collection

In this study, we performed welding experiments using FSW to join similar and dissimilar materials. Four sets of welding experiments were conducted, as detailed in Fig. 1 and Table 1. The first set (Set 1) involved welding aluminum sheets using a cylindrical tool. In this case, the aluminum sheets were of 6061 grade and had the same thickness. These sheets were butt-welded with the cylindrical tool, the dimensions of which are depicted in Fig. 1. The second

(Set 2), third (Set 3), and fourth (Set 4) sets involved lap welding aluminum and steel sheets. In these sets, the aluminum sheets were of 6061 grade, while the steel sheets were AISI304. Furthermore, these sheets had similar thicknesses in Sets 2 and 3 but were welded using the cylindrical tool in Set 2 and the tapered conical tool in Set 3. Set 4 consisted of sheets having dissimilar thicknesses, which were welded using the cylindrical tool.

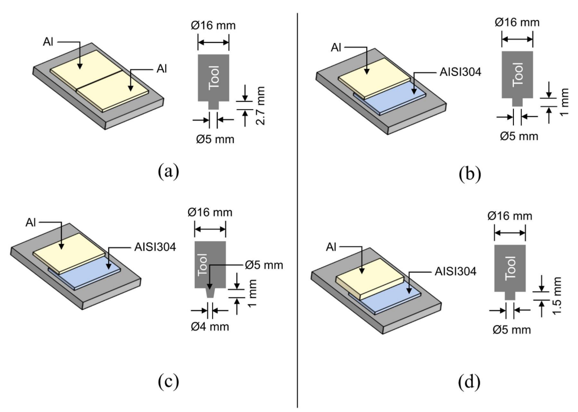



**Fig. 1 Welding experiments: (a) Set 1, (b) Set 2, (c) Set 3, and (d) Set 4**

**Table 1 Experimental set**

| Experiments | Welding configuration | Materials | Tool | Samples welded |
|---|---|---|---|---|
| Set 1 | Similar materials | Al to Al | CY | 42 |
| Set 2 | Dissimilar materials | Al to Steel *(Same thickness)* | CY | 12 |
| Set 3 | | | TC | 12 |
| Set 4 | | Al to Steel *(Dissimilar thickness)* | CY | 12 |

(Al: Aluminum, CY: Cylindrical, TC: Tapered conical)

Throughout the experiments, welding data was collected on the vertical force acting on the welding tool, the torque acting on the spindle, and the total power used during welding. These measurements were sampled at a frequency of 10 Hz.

## Weld Strength

Each welded sample was tested to measure the strength of the weld. This strength is called the ultimate tensile strength and it indicates how close the joint's strength is to the strength of the workpieces. A test specimen was taken from each welded sample to measure its strength. The samples from Set 1 were cut as per the ASTM E8 procedure, which is typically used for butt-welded aluminum sheets. The samples from the other sets were prepared in a rectangular shape since they were lap-welded.

# Methodology

Fig. 2 depicts a schematic of the methodology of this study, which involves welding experiments, analyzing data, creating models, and selecting the best approach for modeling.

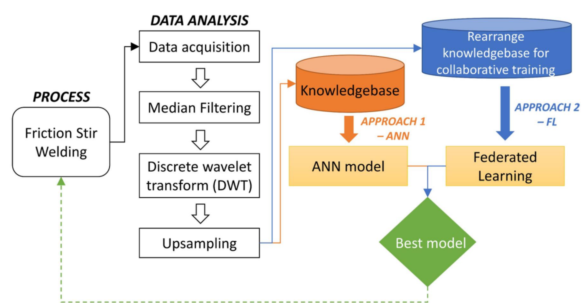



**Fig. 2 Methodology**

## Data Analysis

Discrete wavelet transform (DWT) was used to analyze the signals and identify the frequencies present in them, as well as to quantify defective welds in comparison to defect-free welds. The DWT operation can be mathematically expressed as:

$$W_{\varphi}(j_0, k) = \frac{1}{\sqrt{M}} \sum_{x} f(x) \varphi_{j_0,k}(x) \qquad \text{Eq. (1)}$$

$$W_{\psi}(j,k) = \frac{1}{\sqrt{M}} \sum_{k} f(x) \psi_{j,k}(x)$$

where, $f(x)$ represents the time-series signal, $\varphi$ is the scaling function, $\psi$ is the wavelet function, and $W_{\varphi}$ and $W_{\psi}$ are the approximation and detail coefficients, respectively. In this study, the time series signals available for analysis include the vertical force, spindle torque, and total power, and they have been processed using DWT. Among the two sets of coefficients, the detail coefficients ($W_{\psi}$) are particularly important as they capture the transient changes in the signal, which can provide insight into the quality of the welded samples. Consequently, the detail coefficients were used in the development of the model.

Three levels of detail coefficients were extracted from the signals, with particular emphasis on the first-level coefficients (D1), which were used to predict the weld strength. A total of 64 data points for force, torque, and power signals were transformed into the wavelet space, resulting in 35 data points for each signal. Prior to applying DWT, a median filter of order 15 was used to filter the 64 data points.

The study also compared the benefits of transformed data over raw data by determining their standard deviations. The standard deviation measures the variation within a set of data points, where a low standard deviation suggests that the values are close to the mean, while a high standard deviation indicates that the values are spread out over a wider range. Mathematically, the standard deviation of a discrete set of data $x_1$, $x_2$,..., $x_n$, with each value having the same probability can be given as:

$$\sigma = \sqrt{\frac{1}{N}[(x_1 - \mu)^2 + (x_2 - \mu)^2 + \ldots + (x_N - \mu)^2]} \quad \text{Eq. (2)}$$
$$\mu = \frac{1}{N}(x_1 + \ldots + x_N)$$

or using the summation notation,

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \quad \text{Eq. (3)}$$
$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Considering the set of values of probabilities, i.e., $x_1$ with a probability $p_1$, $x_2$ with probability $p_2$, and so on, the standard deviation of the dataset can be expressed as:

$$\sigma = \sqrt{\sum_{i=1}^{N} p_i(x_i - \mu)^2} \quad \text{Eq. (4)}$$
$$\mu = \sum_{i=1}^{N} p_i x_i$$

Following signal transformation, the transformed data were combined to form a knowledgebase comprising 105 (35×3) data points. The dataset was then upsampled using the spline interpolation to increase the number of data points. This technique involves using neighboring data points to increase the density of the dataset.

## Approach 1: ANN

The ANN model was created to forecast the strength of welds using the multi-featured vector containing D1 coefficients for force, torque, and power signals. The ANN was trained using the backpropagation algorithm applied using the gradient descent method. This method optimizes the error function by iteratively adjusting weights in each epoch. During the modeling process, the momentum was set to 0.9, and the hidden units in each hidden layer were determined using the genetic algorithm (GA). Fig. 3 presents a schematic of the architecture of the optimized ANN.

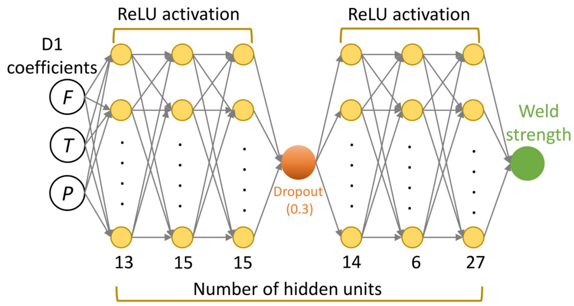



**Fig. 3 Schematic of the optimized ANN model**

The genetic algorithm provides a nearly optimal solution for complex problems. It involved an arbitrary number of neurons in the hidden layers with an upper limit in the initial generation, ultimately resulting in the optimized architecture shown in Fig. 2. The mean squared error (MSE) was chosen as the fitness function, and the Rectified Linear Unit (ReLU) was used as the activation function.

The optimized ANN was trained using the DWT feature vector of size $N \times 105$, where $N$ is the number of samples. The number of features was divided into D1 coefficients of force, torque, and power data, each having 35 features concatenated sequentially. The output aimed to predict the weld strength.

## Approach 2: Federated Learning

The FL approach is a method that allows multiple devices to collaboratively train a prediction model. Each device has its own local data and uses it to update the model independently. These devices create a federated collection, collecting data from individual devices and sharing it with others in the federation [21]. The workflow of this approach includes the following steps: (a) a device downloads the current model and trains it using the data on the device, (b) the device summarizes all the changes (new weights and biases) in a small update, (c) this update is sent to the cloud using encryption techniques for communication, (d) the updates from multiple devices are averaged out, improving the old model to a better global model, and (e) the individual updates are discarded.

The FL approach is schematically shown in Fig. 4. The process begins with a copy of the model localized on each device. In Fig. 4, "A" represents the device that trains a model locally and saves the updates. These model updates are then aggregated from many devices, which highlighted as "B". Finally, these updates are sent to the central server, containing the global model "C" that replaces the old model. The latest model, superior to the individual models, is then sent back to the individual devices.

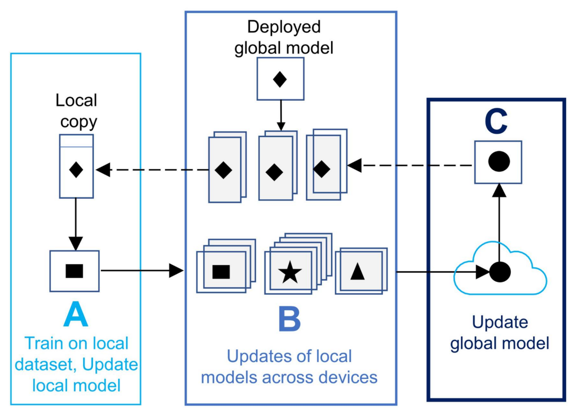



**Fig. 4 Implementing FL**

In this way, the FL approach trains a global and improved model by aggregating the updates of local models trained across devices. Furthermore, this approach eliminates the need for storing client-side training data in the cloud.

In this study, the FL approach was utilized to develop the universal model for diverse welding experiments. In the traditional model development approach, individual

ML or DL models, say $M_i$, will be developed locally using the respective data of $N$ devices, which can be represented as $\{F_1, F_2 .... F_N\}$. In this case, the dataset can be expressed as:

$$D = D_1 \cup D_2 \cup \ldots D_N \qquad \text{Eq. (5)}$$

which will be used to train the model say $M_{sum}$. However, in the FL approach, each device sends model parameters to the centralized server, which aggregates and updates them to form a global model $M_{fed}$. Additionally, the accuracy rate of $M_{fed}$, taken as $V_{fed}$ should not be much greater than the performance of $M_{sum}$, which has $V_{sum}$ accuracy. Let $\delta$ be a non-negative real number if,

$$|V_{fed} - V_{sum}| < \delta \qquad \text{Eq. (6)}$$

So, it can be said that the FL algorithm has $\delta$ - accuracy loss.

Fig. 5 illustrates the implementation of the FL approach. Two devices, $F_1$ and $F_2$, were considered to present the development of the universal model. This model was deployed as a local model for the devices at the beginning of round 1 of training. Training was carried out for 12 rounds, and the dataset was divided between the devices. In each round, a share of data was extracted by each device from its dataset, and the local models were trained. The weights and biases were then aggregated, averaged, and set to the global model for testing against the global testing dataset to determine accuracy.

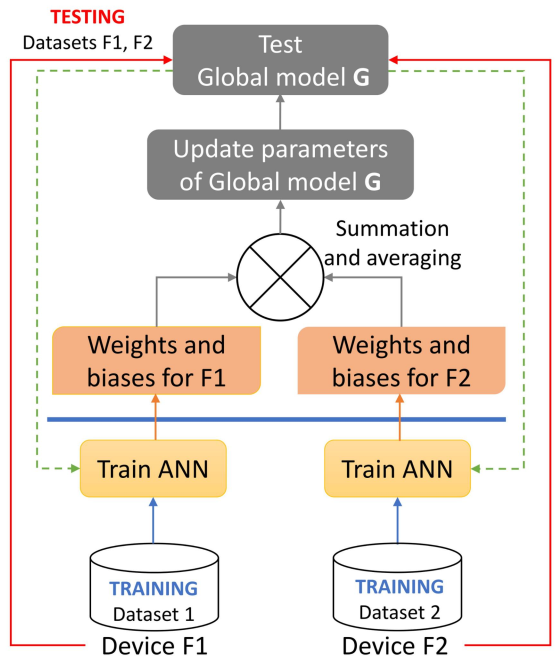



**Fig. 5 Schematic of FL approach**

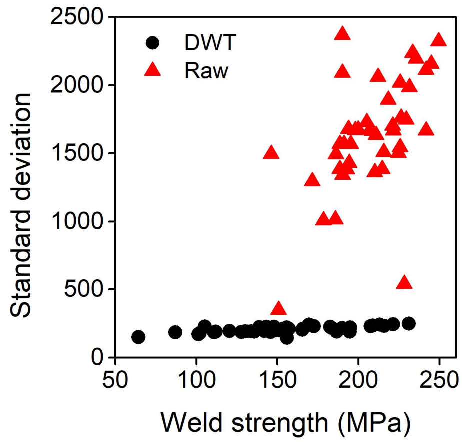
Further, the global model had the same number of hidden neurons and hidden layers as obtained for the optimized ANN model. The first welding experiment set (aluminum-to-aluminum) was used, and training was carried out up to 12 rounds. The data was divided in an IID fashion for comparison with other models.

## Result and Discussion

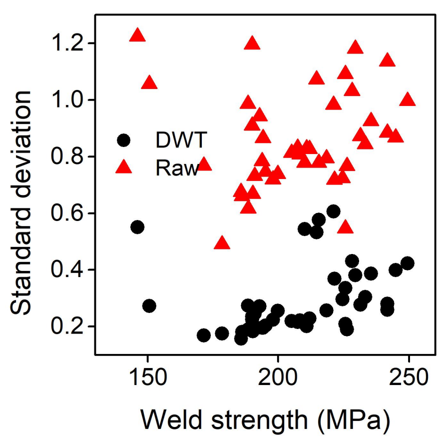
This section presents the results achieved by applying signal processing, model training, and comparative analyses of the ANN and FL approaches.

### Standard Deviation Trends and Validity of using DWT

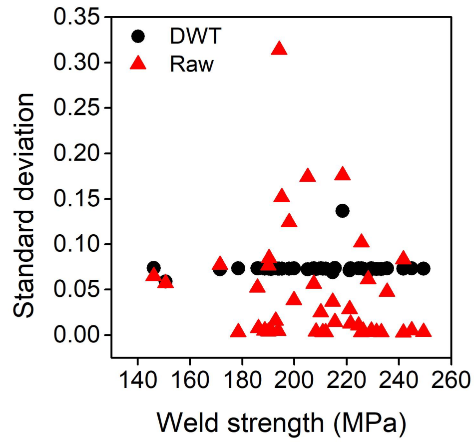
Figs. 6 (a, b, and c) illustrate that the standard deviations of the raw data points for the corresponding weld strength values are significantly higher and more scattered compared to the standard deviations of the DWT coefficients. Additionally, a linear trend is observed for the DWT data of the force and torque signals. However, the standard deviation remains relatively constant for the power signal as weld strength values increase. These trends are not easily discernible from the raw signals due to noise, which is identified as outliers.




**(a)**



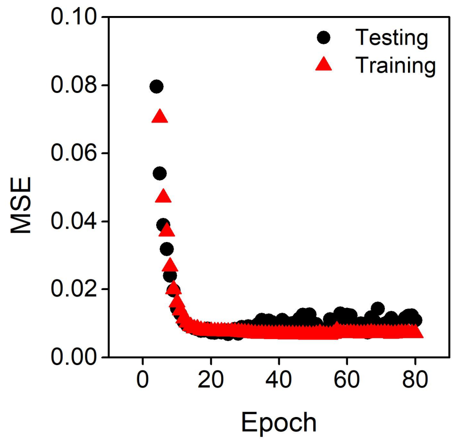

**(b)**



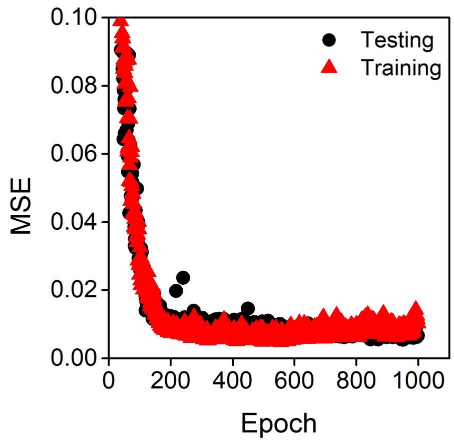

**(c)**

**Fig. 6 Standard deviation of raw and processed data for (a) force, (b) torque, and (c) power**

The trend in the standard deviation data shows that the raw signals contain a lot of noise compared to the DWT data. This observation was further supported when separate ANN models were trained and tested with raw and DWT data. Figs. 7 (a and b) show the mean squared error between actual and predicted weld strength values corresponding to increasing epochs for raw and DWT data, respectively. The raw data has a MSE loss of 0.54865%, whereas the DWT data has a loss of 0.86798%. This observation indicates the necessity of using processed data to reduce the scattering and standard deviation and achieve improved model accuracy.



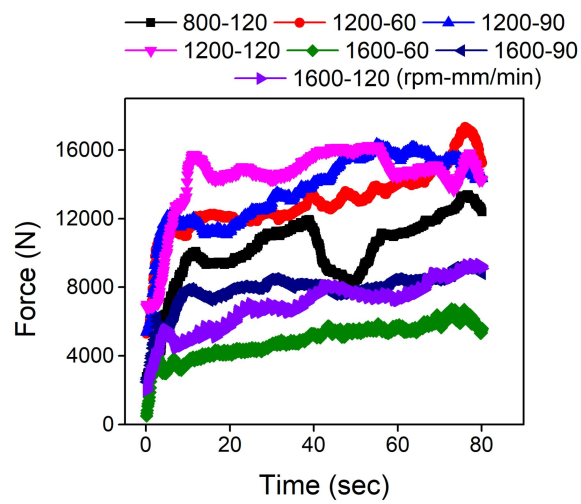

**(a)**



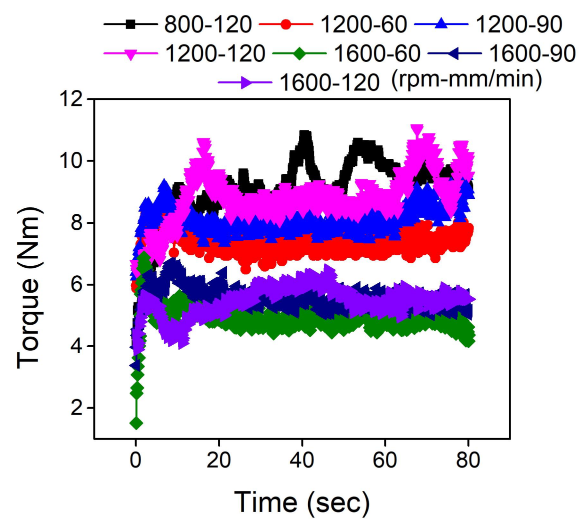

**(b)**

**Fig. 7 MSE of ANN model developed using (a) raw data and (b) processed data**

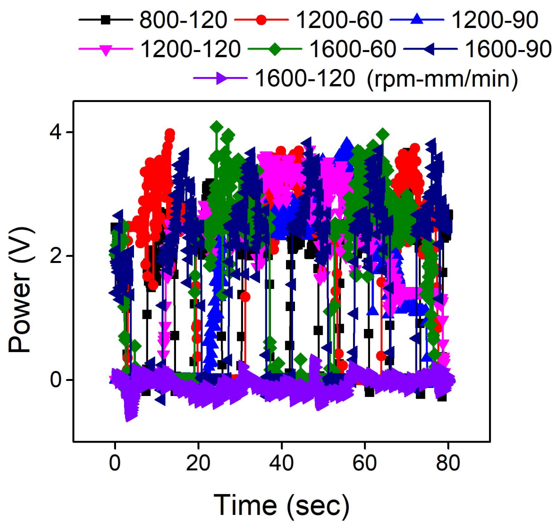
It is important to note that while the sampling frequency was consistent for all signals and experimental sets, the length of the signals may vary due to the welding speed used during the experiments. This speed is determined based on the combination of materials to be welded, type of tool configuration, and the rotational speed of the spindle. Therefore, signals were upsampled to ensure equivalent lengths across all experimental sets. This allowed for identical distributions of the data and model training. A spline interpolation was applied to bound the data points within the set limits. Figs. 8 (a, b, and c) show the upsampled signals of force, torque, and power, resulting in the same length for different combinations of welding parameters.




**(a)**




**(b)**



(c)

**Fig. 8 Upsampled signals (a) force, (b) torque, and (c) power**

## ANN model for predicting weld strength

The first experimental set, aluminum to aluminum, was used to obtain the optimized architecture of the ANN model. Various hyperparameters were tuned, resulting in the best possible configuration for predicting weld strength. Tables 2 and 3 list the MSE and MAE values determined between the actual and predicted weld strength values for different epochs using the Adam optimizer.

**Table 2 MSE of ANN model**

| Epoch | Training | Testing |
|---|---|---|
| 10 | 58.56 | 55.47 |
| 50 | 15.55 | 25.49 |
| 100 | 6.54 | 10.38 |
| 500 | 3.59 | 8.16 |
| **1000** | **1.37** | **8.11** |
| 2000 | 1.49 | 9.25 |

**Table 3 MAE of ANN model**

| Epoch | Training | Testing |
|---|---|---|
| 10 | 1.9 | 12.56 |
| 50 | 1.76 | 12.97 |
| 100 | 1.54 | 12.71 |
| 500 | 1.26 | 12.82 |
| 1000 | 0.88 | 12.74 |
| **2000** | **0.56** | **10.44** |

The best hyperparameter choice from Tables 2 and 3 is 1000 epochs, which resulted in an average training error of 1.37% and a testing error of 8.11% in terms of MSE. These results were obtained using an optimized learning rate of 0.001 and momentum of 0.9 with the Adam optimizer. This configured ANN model was then used to train for the remaining experimental tests, and the results are listed in Table 4.

**Table 4 ANN modeling results for different experimental sets**

| Experiments | Welding configuration | MSE | |
|---|---|---|---|
| | | Training | Testing |
| Set 2 | Al to Steel *(same thickness, CY)* | 1.8206 | 9.6191 |
| Set 3 | Al to Steel *(same thickness, TC)* | 11.2726 | 14.2683 |
| Set 4 | Al to Steel *(dissimilar thickness, CY)* | 9.5496 | 12.6059 |

## Federated Learning for predicting weld strength

Similar to the ANN modeling approach, the aluminum to aluminum experimental set was used to develop the universal model in the FL approach. This model was trained for 12 rounds with the three feature vectors considering two local devices. Fig. 9 shows the MSE determined between the actual and predicted weld strength values for different epochs and rounds for the FL approach. The MSE values converge with increasing rounds, indicating the improvement of the model with increasing epochs, allowing the global model to converge faster even with fewer data points.

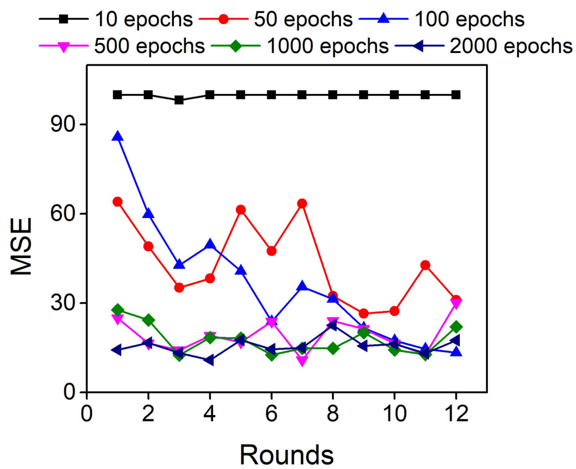



**Fig. 9 MSE of FL approach with increasing epochs and rounds**

Tables 5, 6, and 7 list the validation information, i.e., the MSE values determined between actual and predicted weld strength values for the remaining experimental sets using the FL approach.

**Table 5 MSE of FL approach for aluminum to steel similar thickness welding set using CY tool**

| E | R1 | R2 | R3 | R4 | R5 | R6 | R7 |
|---|---|---|---|---|---|---|---|
| 10 | 176.1 | 100 | 100 | 100 | 100 | 100 | 100 |
| 50 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 100 | 44.9 | 42.84 | 27.39 | 3.95 | 9.70 | 22.17 | 4.84 |
| **500** | 11.3 | 24.11 | 26.81 | **1.86** | 7.23 | 16.07 | 6.612 |
| 1000 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 2000 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

**Table 6 MSE of FL approach for aluminum to steel similar thickness welding set using TC tool**

| E | R1 | R2 | R3 | R4 | R5 | R6 | R7 |
|---|---|---|---|---|---|---|---|
| 10 | 100 | 100 | 100 | 100 | 100 | 100 | 55.61 |
| 50 | 100 | 100 | 100 | 42.69 | 100 | 67.59 | 34.12 |
| **100** | 100 | 100 | 100 | 23.52 | 51.45 | 17.80 | **12.60** |
| 500 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 1000 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 2000 | 20.16 | 21.19 | 25.19 | 18.88 | 23.61 | 17.96 | 23.63 |

**Table 7 MSE of FL approach for aluminum to steel dissimilar thickness welding set using CY tool**

| E | R1 | R2 | R3 | R4 | R5 | R6 | R7 |
|---|---|---|---|---|---|---|---|
| 10 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 50 | 121.39 | 100 | 95.65 | 100 | 11.96 | 22.55 | 17.45 |
| 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 500 | 11.72 | 11.47 | 16.19 | 16.03 | 21.53 | 16.14 | 17.08 |
| 1000 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| **2000** | 14.35 | **9.43** | 10.01 | 16.50 | 20.94 | 13.68 | 12.60 |

Table 8 shows the time complexity obtained with the FL approach for each round. Note that this is determined based on local models, as the FL approach uses data to learn in a decentralized way. Therefore, the trained global model does not reflect the time taken to send the weights and biases to the central server. All the training and testing were conducted on the same machine.

**Table 8 Time complexity obtained with the FL approach with respect to training round (in seconds)**

| Experimental set 1: Al to Al similar thickness CY tool | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| R1 | R2 | R3 | R4 | R5 | R6 | R7 | R8 | R9 | R10 | R11 | R12 |
| 7.91 | 7.91 | 8.00 | 8.00 | 7.96 | 8.46 | 8.01 | 8.06 | 7.95 | 7.99 | 7.95 | 8.12 |
| Experimental set 2: Al to Steel similar thickness CY tool | | | | | | | | | | | |
| 8.61 | 8.70 | 8.27 | 8.25 | 8.21 | 8.58 | 8.27 | -- | -- | -- | -- | -- |
| Experimental set 3: Al to Steel similar thickness TC tool | | | | | | | | | | | |
| 8.27 | 8.55 | 8.24 | 8.19 | 8.17 | 8.30 | 8.36 | -- | -- | -- | -- | -- |
| Experimental set 4: Al to Steel dissimilar thickness CY tool | | | | | | | | | | | |
| 7.87 | 7.83 | 7.82 | 7.81 | 7.85 | 7.98 | 7.93 | -- | -- | -- | -- | -- |

## Performance Comparison of ANN and FL approaches

Fig. 10 shows the mean absolute error (MAE) values for four experimental sets using both ANN and FL approaches. It is clear that the FL approach provides more accurate predictions of weld strength compared to the ANN model. However, when considering cross-platform decentralized training of ML models, the ANN model may struggle with issues of data privacy and bandwidth for sending data to a centralized server. In such a case, the FL model would be more effective. This study established the FL approach and demonstrated that it yields more accurate predictions than the ANN model.

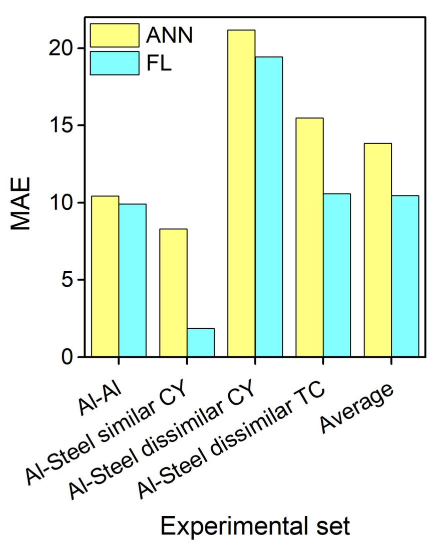



**Fig. 10 MSE for ANN and FL approaches**

Fig. 11 compares the ANN and FL approaches with respect to the epochs required for modeling weld strength prediction. This comparison is shown for the aluminum to steel welding set using materials of similar thickness and CY tool. It is evident that the global model developed with the FL approach converges faster than the traditional ANN, even with a low amount of data in a cross-platform scenario. The FL approach also provides quick and more accurate predictions. In multiple training rounds within the FL approach, in the 4th training round, the global model demonstrates better accuracy despite having less training data than the ANN model. Unlike the ANN model, which uses the entire dataset for a set number of epochs, the FL approach trains and tests models using data chunks. This observation further supports the effectiveness of the FL approach in developing a global model for weld strength prediction over using ANN.

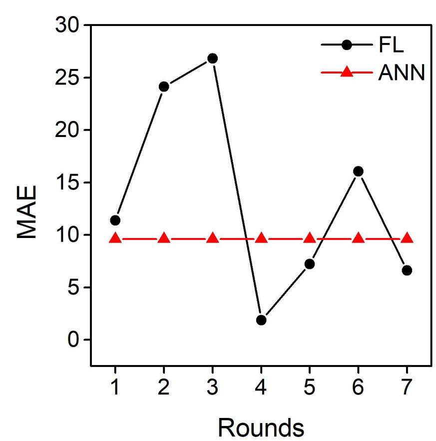



**Fig. 11 MAE for FL (500 epochs) and ANN (1000 epochs) for aluminum to steel similar thickness welding set using CY tool**

One limitation of applying the FL approach to develop a global model is the time required for modeling. Fig. 12 shows the total time taken to train the ANN and global model using the same amount of data. It is evident that the FL approach takes the longest time to train the global model because each round has to train the same number of epochs on two local models, while the ANN trains in a single session.

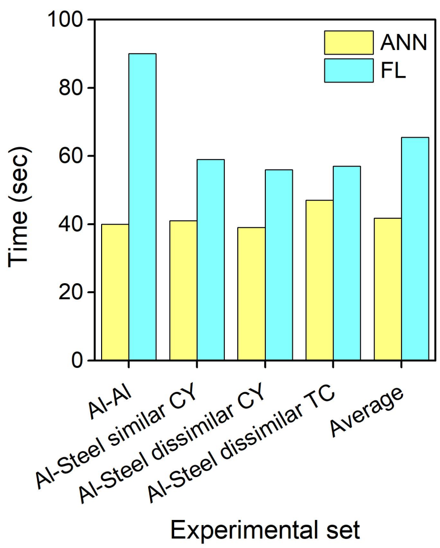



**Fig. 12 Time taken for the training ANN and global models**

This additional time taken by the FL approach could be seen as a trade-off for obtaining a more converged, reliable, and comprehensive global model compared to multiple ANN models, each for one device. Nonetheless, the global model developed using the FL approach is superior to the ANN. Therefore, the FL approach is a better choice in terms of accuracy, as evident from the results. It also remains a better choice from a time perspective, as the training of the global model does not occur in real-time computing and prediction. It occurs once all the data has been collected after a welding process is completed.

## Conclusions

This study utilized the FL approach to create a global model for predicting weld strength in FSW. This approach offers the advantage of a universal system that can accommodate variations in welding experiments across materials, tooling, and welding configurations. In addition to the global model, an ANN model was also developed for comparison. It was found that while the ANN model accurately predicts weld quality, it struggles when used with a new machine. On the other hand, the global model effectively predicts weld strength by sharing data across different experimental sets, making it more suitable for industrial use. Moreover, in situations involving variations in experimental sets, cloud services for data collection, storage, and analysis may be necessary, which can increase computational resources and overall costs. However, the FL approach allows data from individual experimental sets to be stored on local machines, while the FL model is updated globally, reducing computational resources.

One challenge that could hinder the FL approach is the communication within the federated network. As the FL approach relies on updates flowing to the central server from its federated networks, communication glitches within the industrial shop floor can affect model training. Another challenge is the collection of welding data involving multiple variants, which can be time-consuming and expensive.

Although this study primarily focused on the performance of the global model, future research will aim to optimize this model. Optimizing the global model poses challenges as it needs to accommodate multiple devices, which may be significantly different. Additionally, the practical application of the FL approach in an industrial shop floor would require establishing a secure connection for real-time monitoring and control, which will also be explored in the future.

## References


[1] A. B. Nassif, M. A. Talib, Q. Nasir, and F. M. Dakalbab, “Machine Learning for Anomaly Detection: A Systematic Review,” *IEEE Access*, vol. 9, pp. 78658–78700, 2021, doi: 10.1109/ACCESS.2021.3083060.

[2] L. Monostori, J. Hornyák, C. Egresits, and Z. J. Viharos, “Soft computing and hybrid AI approaches to intelligent manufacturing,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, 1998, p. 765. doi: 10.1007/3-540-64574-8_463.

[3] S. Thudumu, P. Branch, J. Jin, and J. Singh, “A comprehensive survey of anomaly detection techniques for high dimensional big data,” *J. Big Data*, vol. 7, no. 1, p. 42, Dec. 2020, doi: 10.1186/s40537-020-00320-x.

[4] S. Trilles, S. S. Hammad, and D. Iskandaryan, “Anomaly detection based on Artificial Intelligence of Things: A Systematic Literature Mapping,” *Internet of Things*, vol. 25, p. 101063, Apr. 2024, doi: 10.1016/j.iot.2024.101063.

[5] S. Wang and J. Zhang, “An Intelligent Process Fault Diagnosis System based on Andrews Plot and Convolutional Neural Network,” *J. Dyn. Monit. Diagnostics*, pp. 127–138, May 2022, doi: 10.37965/jdmd.2022.67.

[6] “Deep Learning and Its 5 Advantages | by Magnimind | Becoming Human: Artificial Intelligence Magazine.”

[7] S. K. Thangarasu, T. Mohanraj, K. Devendran, M. Rajalakshmi, S. Chowdhury, and S. Gurusamy, “Prediction of Flank Wear during Turning of EN8 Steel with Cutting Force Signals Using a Deep Learning Approach,” *Math. Probl. Eng.*, vol. 2023, pp. 1–12, Feb. 2023, doi: 10.1155/2023/5401372.

[8] M.-H. Tsai, J.-N. Lee, H.-D. Tsai, M.-J. Shie, T.-L. Hsu, and H.-S. Chen, “Applying a Neural Network to Predict Surface Roughness and Machining Accuracy in the Milling of SUS304,” *Electronics*, vol. 12, no. 4, p. 981, Feb. 2023, doi: 10.3390/electronics12040981.

[9] B. Savkovic, P. Kovac, D. Rodic, B. Strbac, and S. Klancnik, “Comparison of artificial neural network, fuzzy logic and genetic algorithm for cutting temperature and surface roughness prediction during the face milling process,” *Adv. Prod. Eng. Manag.*, vol. 15, no. 2, pp. 137–150, Jun. 2020, doi: 10.14743/apem2020.2.354.

[10] C. Wu, C. Wang, and J.-W. Kim, “Welding sequence optimization to reduce welding distortion based on coupled artificial neural network and swarm intelligence algorithm,” *Eng. Appl. Artif. Intell.*, vol. 114, p. 105142, Sep. 2022, doi: 10.1016/j.engappai.2022.105142.

[11] O. Mypati, A. Mukherjee, D. Mishra, S. K. Pal, P. P. Chakrabarti, and A. Pal, “A critical review on applications of artificial intelligence in manufacturing,” *Artif. Intell. Rev.*, vol. 56, no. S1, pp. 661–768, Oct. 2023, doi: 10.1007/s10462-023-10535-y.

[12] R. Y. Zhong, X. Xu, E. Klotz, and S. T. Newman, “Intelligent

Manufacturing in the Context of Industry 4.0: A Review,” *Engineering*, 2017, doi: 10.1016/J.ENG.2017.05.015.

[13] K. Balachandar, R. Jegadeeshwaran, and D. Gandhikumar, “Condition monitoring of FSW tool using vibration analysis-A machine learning approach,” *Mater. Today Proc.*, vol. 27, pp. 2970–2974, 2019, doi: 10.1016/j.matpr.2020.04.903.

[14] B. Das, S. Pal, and S. Bag, “Weld quality prediction in friction stir welding using wavelet analysis,” *Int. J. Adv. Manuf. Technol.*, vol. 89, no. 1–4, pp. 711–725, 2017, doi: 10.1007/s00170-016-9140-0.

[15] A. K. Lakshminarayanan and V. Balasubramanian, “Comparison of RSM with ANN in predicting tensile strength of friction stir welded AA7039 aluminium alloy joints,” *Trans. Nonferrous Met. Soc. China*, vol. 19, no. 1, pp. 9–18, Feb. 2009, doi: 10.1016/S1003-6326(08)60221-6.

[16] T. Nasir, M. Asmaela, Q. Zeeshan, and D. Solyali, “Applications of machine learning to friction stir welding process optimization,” *J. Kejuruter.*, vol. 32, no. 2, pp. 171–186, 2020.

[17] S. Verma, J. P. Misra, J. Singh, U. Batra, and Y. Kumar, “Prediction of tensile behavior of FS welded AA7039 using machine learning,” *Mater. Today Commun.*, vol. 26, p. 101933, 2021, doi: 10.1016/j.mtcomm.2020.101933.

[18] D. Mishra *et al.*, “Real time monitoring and control of friction stir welding process using multiple sensors,” *CIRP J. Manuf. Sci. Technol.*, no. 2019, Apr. 2020, doi: 10.1016/j.cirpj.2020.03.004.

[19] D. Mishra *et al.*, “Industry 4.0 Application in Manufacturing for Real-Time Monitoring and Control,” *J. Dyn. Monit. Diagnostics*, pp. 176–187, Sep. 2022, doi: 10.37965/jdmd.2022.118.

[20] A. H. Elsheikh, “Applications of machine learning in friction stir welding: Prediction of joint properties, real-time control and tool failure diagnosis,” *Eng. Appl. Artif. Intell.*, vol. 121, p. 105961, May 2023, doi: 10.1016/j.engappai.2023.105961.

[21] “Google AI Blog: Federated Learning: Collaborative Machine Learning without Centralized Training Data.”