End-to-end online quality prediction for ultrasonic metal welding using sensor fusion and deep learning

Yulun Wu[a,1], Yuquan Meng[a,1], Chenhui Shao[a,*]

[a] *Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*

Abstract

In industrial-scale production applications of ultrasonic metal welding (UMW), there is a strong need for predicting joint quality quickly, reliably, and non-destructively. State-of-the-art quality assessment methods such as destructive tensile testing and binary quality classification cannot meet such requirements. This paper develops a novel end-to-end online quality prediction method for UMW based on sensor fusion and deep learning. This method first preprocesses 1-dimensional signals from multiple sensors including an acoustic emission sensor, a linear variable differential transformer, and a microphone, and transforms them to 2-dimensional images using wavelet transform. Then, these images are fed into ResNet20, which is a 20-layer convolutional neural network, to automatically generate feature maps and predict joint strength. The proposed method offers important advantages compared to state-of-the-art approaches, including automatic feature generation and good robustness to UMW tool conditions. The effectiveness of the developed method is demonstrated using real-world data generated from an UMW process with four different tool conditions. Additionally, we propose three feature fusion strategies (early fusion, middle fusion, and late fusion) and present a comparative case study to compare their performance. It is found that the late fusion strategy achieves the best prediction performance.

[*] Corresponding author

Email addresses: `yulunwu4@illinois.edu` (Yulun Wu), `yuquanm2@illinois.edu` (Yuquan Meng), `chshao@illinois.edu` (Chenhui Shao)

[1] Yulun Wu and Yuquan Meng have contributed equally to the work.

Towards interpretability and explainability in deep learning, we perform a correlation analysis to reveal the connection between ResNet-generated features and features that are manually extracted based on UMW process physics. It is shown that many manual features are strongly correlated with ResNet features, proving that ResNet is able to resemble physical knowledge. The proposed online quality prediction method is readily applicable to industrial-scale UMW processes to enable accurate online quality prediction.

Keywords: ultrasonic metal welding, quality prediction, deep learning, ResNet, sensor fusion, interpretability

2010 MSC: 00-01, 99-00

1. Introduction

Ultrasonic metal welding (UMW) is a solid-state welding process in which multiple sheets of either similar or dissimilar metal materials are joined using high-frequency vibrations in plane with the interfaces under pressure [1, 2]. A typical UMW system is comprised of controller, transducer, booster, horn, and anvil. During an UMW cycle, high-frequency vibration is applied to clamped metals, which undergo material softening, plastic deformation, and fatigue crack formation, and finally a joint forms [2, 3]. Because of its unique advantages including the ability to join dissimilar materials, energy efficiency, environmental friendliness, and short process cycles, UMW is an important joining technology with wide industrial applications such as electric vehicle battery assembly [1, 4] and automotive body joining [2, 5].

In industrial-scale production applications, the consistency and quality of UMW need to be closely monitored [6–9]. Several key process parameters such as welding time, amplitude, and pressure have been shown to significantly influence the joint quality. Therefore, process optimization is usually performed to identify the optimal parameter combination [10, 11]. However, even if those process parameters are fixed in mass production, the joint quality may still vary substantially due to uncontrollable factors such as material surface condition

[1, 6] and tool condition [12–14]. Lee et al. demonstrated that the surface condition of welding specimens significantly influences joint quality [1]. Drawing on this finding, Nong et al. developed a control method to detect surface contamination and accordingly adjust welding pressure, which was shown to greatly improve joint quality [6]. Tool degradation is a major concern in industrial UMW because tool wear rate is high [12, 13] and tool conditions substantially impact joint quality [14]. It was reported that tool maintenance constitutes a major part of production costs in UMW [13]. Further, accurately monitoring the changes in tool condition is challenging because of the complicated tool surface geometry [13, 15]. Consequently, it is highly desirable to develop a quality prediction system that can adequately account for these uncontrollable factors and accurately predict the variability in joint quality.

Most existing works on UMW quality prediction fall into three categories, namely, physics-based simulations [16–18], data-driven response surface modeling [10, 11, 19–21], and quality classification using sensing signals [1, 7, 9, 22, 23], which are briefly reviewed as follows.

Physics-based simulations typically employ finite element (FE) models to investigate the joining mechanism or predict joint performance. For example, Xi et al. [16] developed FE models to predict the mechanical performance of UMW. Zhou et al. [17] conducted FE analysis to predict weld strength using the cohesive zone method. Shen et al. [18] developed 3D FE models to simulate the complex material response subject to UMW processing conditions and offer important insights into the design of UMW applications. These simulation models provide a good physical understanding of the relationship between the welding processes and joint quality. However, FE simulations are time-consuming, and thus cannot meet the requirement of online monitoring. Furthermore, process variability induced by uncontrollable factors cannot be well captured by such methods.

The second type of approach builds statistical or machine learning-based response surface models to characterize the relationship between process parameters and joint quality, which can be used for process optimization. For instance, Kim et al. [19] adopted cubic functions to establish a response surface for the peel strength using welding time and pressure as inputs in UMW of Cu and Ni-plated Cu sheets. Using artificial neural network and adaptive neuro-fuzzy inference system, regression models were developed in [20] to predict the strength for Al-Cu joints. Meng et al. [10] used machine learning models to study the impact of amplitude and weld time on peel and shear strengths, based on which a multi-objective optimization method was subsequently developed. A hybrid machine learning model was developed in [21] to predict the quality of Al-Al UMW joints from welding parameters and peak power. Yang et al. [11] developed a hybrid multi-task learning method to efficiently model the response surfaces of UMW with different material combinations. Those methods are deterministic and based on the assumption that the welding quality is mainly decided by the input welding parameters but generally ignore the impact of uncontrollable factors. The variability in joint quality when identical process parameters are used cannot be accounted for.

The third type of method aims to classify joint quality online and non-destructively by exploring the relationship between online sensing signals and the joint quality. For example, Shao et al. [22] developed an algorithm for feature selection and parameter tuning in quality classification. Lee et al. [1] determined the correlation between signal features and joint performance. In [7], the relationships between the displacement of the sonotrode, the plastic deformation of materials, and the joint quality were established for process monitoring purpose. Guo et al. [23] integrated Shewhart-type control chart and M-distance approach to detect defective welds online. Shi et al. [9] proposed a monitoring method to detect multiple abnormal welding conditions using power signals. Nevertheless, most of these works focus on classification but cannot predict joint strength, i.e., regression. Further, extensive feature engineering is often required in these methods, which heavily relies on a good understanding of the process physics and may be tedious.

Most recently, some studies were focused on quantitatively predicting UMW joint strength [8, 24]. Schwarz et al. [24] extracted quality monitoring features from multiple sensors and then developed linear and multi-layer perceptron regression models to predict the tensile shear strength of Cu-sheet welds. Improved prediction accuracy was achieved compared to standard polynomial regression models. Meng and Shao [8] developed a physics-informed ensemble learning framework for UMW joint strength prediction. A feature extraction procedure based on discrete wavelet transform was proposed to alleviate tedious feature engineering efforts. Nevertheless, these existing methods still require a good understanding of the process physics and cannot offer the capability of automatic, end-to-end online joint strength prediction.

To overcome the aforementioned challenges, we establish an end-to-end online quality prediction method using deep learning and sensor fusion. Here, "end-to-end" refers to capabilities of bypassing complicated intermediate procedures (e.g., feature extraction, feature selection, model selection, model parameter tuning) and predicting outputs (joint strength) from inputs (sensing signals) directly. In this method, 1-dimensional (1D) signals from acoustic emission (AE) sensor, linear variable differential transformer (LVDT), and microphone are first processed to filter out redundant information. Then, wavelet transform is applied to convert 1D signals to 2-dimensional (2D) images. We use ResNet to process the 2D images and build a regression model in an end-to-end fashion. Specifically, ResNet automatically generates features from 2D images and subsequently uses the features to predict the joint strength. The effectiveness of the developed method is demonstrated using real-world UMW data. We also propose three sensor fusion strategies including early fusion, middle fusion, and late fusion and compare their performance.

The contributions of this paper are summarized as follows:

(1) A novel end-to-end online prediction model is developed for UMW based on deep learning that offers various benefits, including superior quality prediction, minimal reliance on prior knowledge of UMW processes (e.g., tool conditions), and not involving tedious data preprocessing and feature engineering.

(2) We propose three sensor fusion strategies, namely, early fusion, middle fusion, and late fusion, that can effectively fuse information from multiple sensing signals for joint strength prediction.

(3) Using experimental UMW data, comprehensive case studies are constructed to compare (1) the proposed prediction method and traditional feature engineering-based methods and (2) three sensor fusion strategies. It is shown that our method achieves higher prediction accuracy. In addition, the late fusion strategy using displacement and sound signals achieves the best prediction performance.

(4) Towards interpretability and explainability in deep learning, we perform a correlation analysis to reveal the connection between ResNet-generated features and manually extracted features. It is found that many manual features are strongly correlated with ResNet features, which demonstrates that deep learning is able to resemble physical knowledge.

The remainder of the paper is organized as follows. Section 2 presents the workflow of the joint strength prediction framework. Section 3 explores the collected UMW data and conducts an exploratory analysis. In Section 4, case studies are reported to demonstrate the effectiveness of the proposed method. Finally, Section 5 concludes the paper and suggests future research directions.

2. End-to-end quality prediction framework

The workflow of the proposed method is shown in Fig. 1. There are mainly three steps. The first step is signal processing including signal truncation, signal down sampling, and wavelet transform. The second step is model training. In this step, a deep learning model is trained to automatically generate feature maps using processed signals obtained in the first step. The final step is to predict joint strength. Sections 2.1 and 2.2 will elaborate the signal processing procedure and the ResNet model used in this study, respectively.

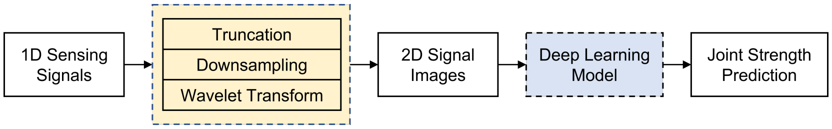

Fig. 1. Workflow of the end-to-end online quality prediction method.

2.1. Data processing

Our proposed algorithm uses signals from three online sensors including AE, LVDT, and microphone as inputs. For each UMW process, one sensing signal consists of more than 200 million data points. Therefore, signal processing is necessary to remove the noise and irrelevant information in the signals. An UMW process consists of multiple stages, some of which (e.g., preparation) do not influence the joint quality significantly [23]. Therefore, we apply signal truncation to extract the signal segment corresponding to the main vibration step of the welding process.

After down sampling, the 1D signals are transformed to 2D images using wavelet transform [25]. Because the joining mechanism of UMW is primarily determined by high-frequency vibration [26], the vibration pattern impacts the formation of joints and ultimately determines the joint quality. On the other hand, different vibration patterns can be sensed by sensing signals. The images generated by wavelet transform reflect the changes in the vibration pattern over time, thereby containing useful information about the joint quality.

The discrete wavelet transform can be obtained by:

$$T_{j,k} = \int_{-\infty}^{+\infty} f(t)\psi_{j,k}(t)dt, \tag{1}$$

$$\psi_{j,k}(t) = \frac{1}{\sqrt{2^k}}\psi\left(\frac{t-j2^k}{2^k}\right), \tag{2}$$

where $\psi(t)$ is mother wavelet function, $f(t)$ is the original 1D signal (e.g., AE, LVDT, microphone) and $T_{j,k}$ is the coefficient for each wavelet, j is the shift parameter, and 2^k controls the scale of the wavelet. By introducing the shift and scale parameters, both the occurring time and frequency information of wave component in raw signals can be revealed. The norm of $T_{j,k}$ reflects the energy contained in wavelet $\psi_{j,k}(t)$.

Fig. 2 shows a wavelet transform example for an AE signal. Fig. 2(b) shows the distribution of energy in the AE signal in time-frequency domains. In Fig. 2(b), each horizontal line represents a component for a certain frequency in the AE signal, while each vertical line shows the energy distribution for each frequency at a certain time. There are several energy concentration points in the frequency spectrum, indicating that the vibration of the corresponding frequency plays a dominant role in that specific time during the welding process. Additionally, it is seen that the energy distribution with respect to frequency changes over time. Because high-frequency vibration is the main process mechanism in UMW, capturing the vibrational information is crucial for joint quality prediction. As such, the wavelet transform-generated image potentially provides important insights into the welding process and can be used to predict the joint strength.

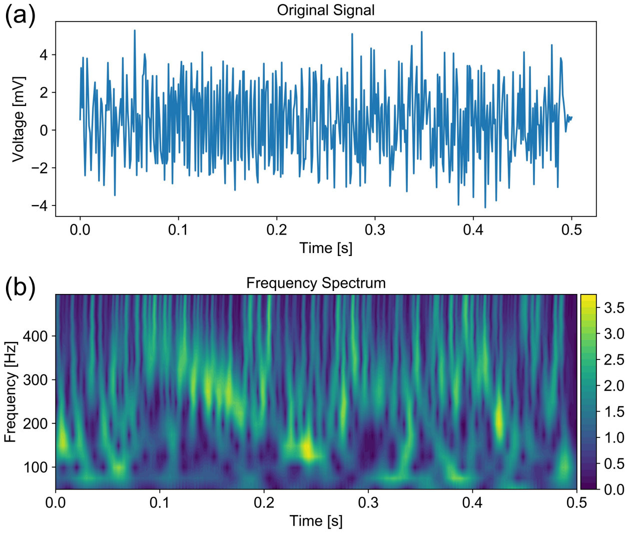

Fig. 2. A wavelet transform example for an AE signal.

Deep learning models are good candidates to generate useful features and find the relationship between the 2D images and the joint strength because of its strong ability to characterize complex nonlinear input-output relationships. They also avoid tedious feature engineering efforts that are required if conventional machine learning models are employed for prediction.

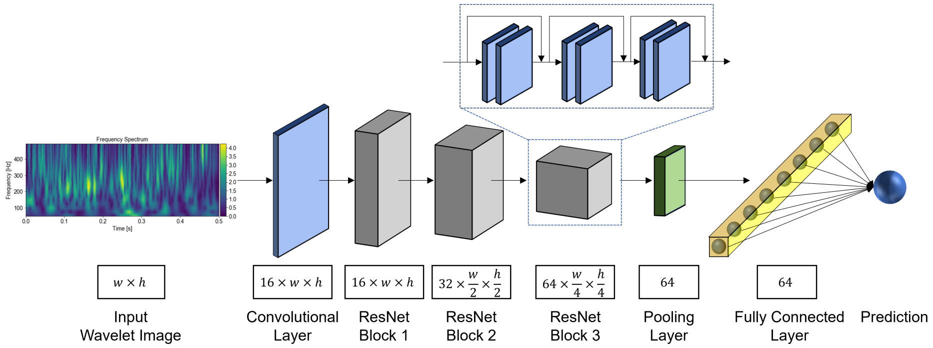

Fig. 3. ResNet20: The architecture of the automatic joint strength prediction framework.

The proposed joint strength prediction framework is shown by Fig. 3. The input of the model is a 2D wavelet image. We use the ResNet structure [27, 28] to accomplish self feature generation. A fully connected layer is used to predict the joint strength based on extracted features. The output of a convolutional layer is:

$$a_{i,j}^{(l)} = f\left(\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} a_{i+u,j+v}^{(l-1)} \cdot k_{rot}(l)_j \cdot \chi(i,j) + b^{(l)}\right), \tag{3}$$

$$\chi(i,j) = \begin{cases} 1, \ 0 \le i,j \le n \\ 0, \ \text{others} \end{cases} \tag{2a}$$

where $a_{i,j}^{(l)}$ is the output of lth convolutional layer, k is an $n \times n$ convolutional kernel, b is the bias and f is the activation function. $a_{i,j}^{(l)}$ can be regarded as weighted $a_{i,j}^{(l-1)}$ and be further regarded as weighted input. All outputs of the convolutional layer can be regarded as features extracted from the input, which is the 2D wavelet image in this method. Therefore, during training process, the deep learning model will autonomously find the pattern in the 2D wavelet images that is related to the joint strength and generate feature maps accordingly. In this way, our proposed model skips the feature generation and selection process that are required by conventional data-driven approaches.

The sizes of the output feature maps are also shown in Fig. 3. w and h stand for the width and height of the input image, respectively. If the size of feature maps is divided by 2, then the number of the convolutional kernel is doubled. A 3 by 3 convolutional kernel is adapted for all convolutional layers. As a result, the original reception field of a convolutional kernel is 3 by 3. With the increase of the depth of the ResNet, the reception field of a kernel also increases accordingly. These kernels are used to generate potential features in their reception fields. The final fully connected layer is used to combine all generated features and generate the prediction of the joint strength.

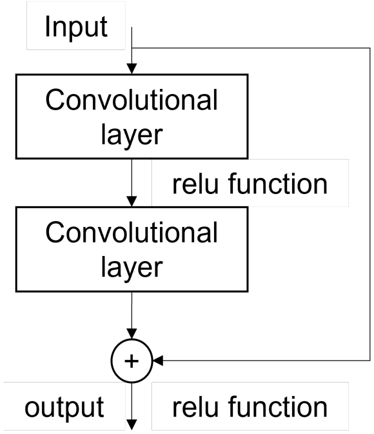

Fig. 4. Structure of a ResNet block unit.

We adopt the ResNet block [27, 28] in our deep learning model as shown in Fig. 4. Every ResNet block contains three residue block units, which can both accelerate the convergence and ensure good performance. The generated feature maps from previous convolutional layers can be used with high-level feature maps, which provide more possible combinations of different features. Three ResNet blocks are used so there are 18 convolutional layers in the ResNet blocks. Including the first convolutional layer and the last fully connected layer, the network has 20 layers in total. As a result, we refer to this architecture as ResNet20 in the rest of this paper.

3. Data exploration

3.1. Experimental setup and preliminary analysis

This research uses the dataset reported in [14] for method validation. The experimental set-up is shown in Fig. 5. Signals from four sensors, namely, power, AE, LVDT, and microphone are collected for each welding cycle. Power signals are directly obtained from the welder controller. Displacement signals are obtained from the LVDT sensor installed in the welder actuator. Microphone signals are captured by GRAS 40PP microphone, which is placed close to the actuator. AE signals are gathered by an external AE sensor R15α, which is attached to the anvil.

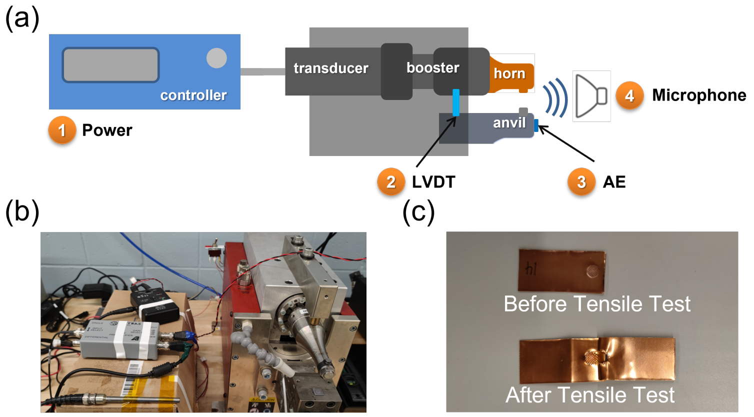

Fig. 5. Experimental setup: (a) schematic of DAQ system [14], (b) photo of the UMW machine and DAQ system, and (c) example photos of weldments before (top) and after tensile test (bottom).

Copper material C110 was used in the welding experiments. The dimension of the copper samples is 50.8 mm × 25.4 mm × 0.2032 mm (length × width × thickness). A Branson Ultra-weld L20 system was used for the welding experiments. 50 repetitive experiments were carried out with four different tool conditions, resulting in 200 welding experiments in total. Table 1 shows the tool conditions used in the experiments. Tool conditions were determined by measuring the tool surfaces with a 3D laser scanning microscope. See [14] for details. Welding parameters were fixed for all welding cycles. Specifically, welding time is 0.5 s, amplitude is 45 µm, and pressure is 45 psi. After welding experiments, all weldments were subject to T-peel test carried out by a universal testing machine MTS 810. During the test, load curves of the testing machine with time were recorded. The joint strength is defined as the maximum load and indicates the joint quality.

Table 1: Tool conditions

Tool Condition	Horn	Anvil
1	New	New
2	New	Worn
3	Worn	New
4	Worn	Worn

Histograms and kernel density estimation plots of joint strength in four tool conditions are shown in Fig. 6. Some key statistics for the joint strength of each tool condition are presented in Table 2. Outliers were removed before the calculations. Obviously, the statistical distributions for four tool conditions are different, confirming that tool conditions influence joint quality substantially. Additionally, it is seen that strong variability exists within each group, which highlights the necessity for online joint quality prediction. One may notice that tool condition 4 has the highest mean strength and relatively small variance. This is because all the welding parameters are predetermined before experiments and not fine-tuned for any specific tool condition. It is thus possible that the parameters are more suitable for tool condition 4 than other tool conditions.

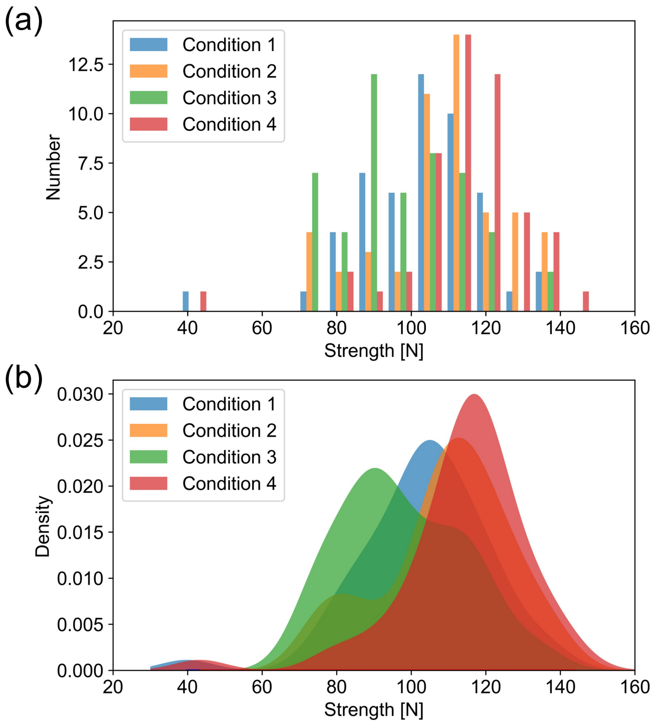

Fig. 6. Distribution of joint strength in four tool conditions: (a) histograms and (b) kernel density estimation plot.

Table 2: Key statistics for joint strength of welded samples produced using four tool conditions.

Tool Condition	Mean (N)	Standard Deviation (N)	Minimum (N)	Maximum (N)
1	104.8	14.7	73.98	138.40
2	109.4	17.0	73.58	140.39
3	97.7	16.4	78.11	143.98
4	115.5	13.5	69.69	134.92

240 It is worth noting that the variation in tool conditions leads to a significant challenge in predicting joint quality. In UMW, tool degradation occurs in the form of knurl height reduction and/or knurl breakage [12, 13, 15]. The knurls

of worn tools have non-uniform height across the tool surfaces [15]. Coupled with the complicated working mechanism of UMW, this non-uniformity leads to significant challenges for studying the influence of tool degradation. Theoretical investigation of how tool degradation influences joint quality is still largely lacking. The literature on joint quality classification and prediction seldom considers this factor. Among the existing works on joint strength prediction, i.e., [8, 24], tool condition is fixed in the experiments in [24]; and a hybrid modeling architecture was developed in [8] to first classify tool conditions and then use a different regression model for each tool condition. In this paper, we aim to develop an end-to-end framework that can account for the influence of tool conditions and perform the regression task automatically.

3.2. Signal visualization and processing

In the UMW experiments, the sampling time for sensors was set to 2 seconds while the welding time was set to 0.5 seconds. To extract the useful segments from the original signals, we use the power signal to find the start and end points for the welding cycle and keep the 0.5-second segment for subsequent analysis. All signal pieces are downsampled using a sampling rate of 1/101.

Wavelet transform is applied to each processed signal to obtain the characteristics of signals in the time-frequency domain. The sensing signals and their wavelet transforms for three representative samples from the low-strength, medium-strength, and high-strength groups are shown in Fig. 7–9. The y axis of frequency spectrum is from 50 Hz to 500 Hz. We perform the wavelet transform for 20 scales to get 19 lines of frequency spectrum. The frequency spectrum matrix is visualized with a color scale. A lighter color means more energy of the frequency at corresponding time in the signal. The dark blue background color indicates limited energy. Three groups of signals are different in both time and frequency domains. Such differences imply that the 2D images generated by wavelet transform contain useful information about joint quality and can be used for predicting joint strength. Yet, due to the high complexity of the welding process, it is very challenging to mine such information and devise effective monitoring features manually. Therefore, we use the ResNet model to find the relationship between the 2D images and joint quality automatically.

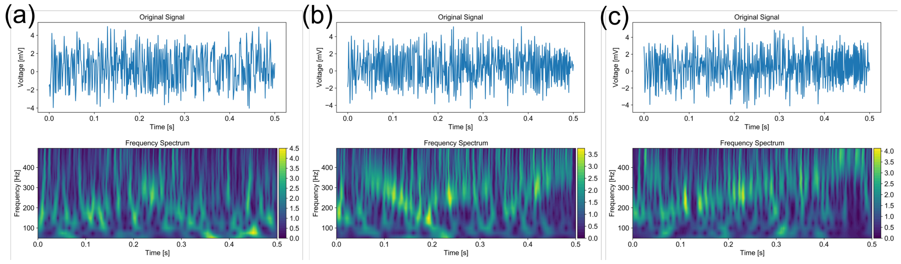

Fig. 7. Downsampled signals and their wavelet transforms for AE signals: (a) low joint strength (69.69 N) with worn horn and worn anvil, (b) medium joint strength (101.44 N) with new horn and new anvil, and (c) high joint strength (134.05 N) with new horn and new anvil.

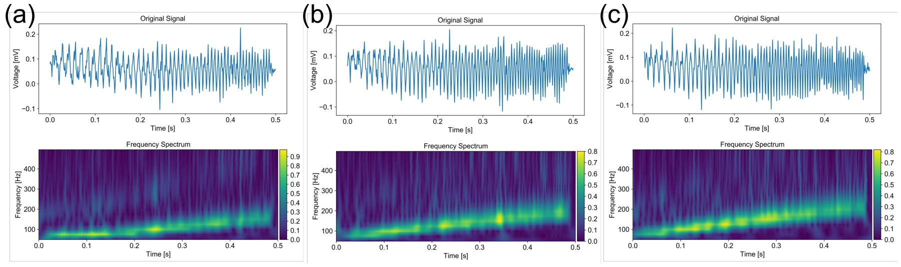

Fig. 8. Downsampled signals and their wavelet transforms for LVDT signals: (a) low joint strength (69.69 N) with worn horn and worn anvil, (b) medium joint strength (101.44 N) with new horn and new anvil, and (c) high joint strength (134.05 N) with new horn and new anvil.

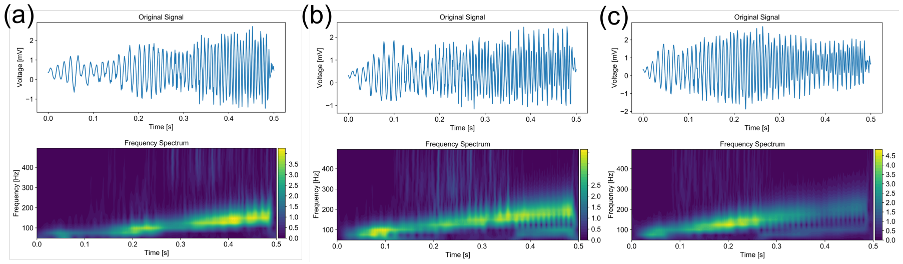

Fig. 9. Downsampled signals and their wavelet transforms for microphone signals: (a) low joint strength (69.69 N) with worn horn and worn anvil, (b) medium joint strength (101.44 N) with new horn and new anvil, and (c) high joint strength (134.05 N) with new horn and new anvil.

4. Results and discussion

This section presents comparative case studies to demonstrate the effectiveness of the proposed method. Implementation details are introduced in Section 4.1. The performance comparison of (2) the proposed method and conventional machine learning models and (3) three sensor fusion strategies are presented in Sections 4.2 and 4.3, respectively.

4.1. Implementation details

A 16 GB Nvidia Tesla P100 GPU is used to train ResNet models. The basic ResNet model contains 100 Kb (822993) trainable variables. For each training process, the model is trained for 200 epochs. For each training epoch, the batch size is set to 20. All ResNet models use the same fixed learning rate of 0.000003. All ResNet models are trained from the scratch. The batch normalization layer is used after each convolutional layer.

In all case studies, we use a 4:1 training-test random split for model performance evaluation and repeat the process five times to avoid contingency. Hyperparameters of all prediction models are carefully tuned and the results from best-performing models are reported. Root mean square error (RMSE) is used to evaluate the prediction performance. The unit of RMSE is N. The means of both training and test RMSEs over five repeats are calculated and compared.

4.2. Performance comparison: single sensor case

We compare our method with conventional approaches in single-sensor cases, i.e., only one type of sensing signal is used for joint strength prediction. Conventional prediction models are built following feature engineering, feature selection, and model training. The details of feature generation and selection are provided in the appendix. The selected features are used to train three classical regression models, i.e., support vector regression (SVR) with radial basis function (RBF) kernel, k-nearest neighbors (KNN), and linear regression (LR). Our method is trained following the workflow of Fig. 1. The comparative results are summarized in Table 3. The lowest average test RMSE (15.42 N) is obtained when SVR and microphone features are used.

On the other hand, the lowest average test RMSE from our method is 15.44 N, which is comparable to that of conventional methods. Moreover, it is important to note that conventional approaches require extensive feature engineering. Over 300 features are manually extracted, and feature selection is then carried out to select the most important features for regression model training. Our method permits a significantly more efficient procedure by using deep learning to generate features automatically.

Table 3 also sheds some light on the efficacy of sensing signals used in the online monitoring system. It is shown that the models using microphone or displacement signals for prediction generally achieve lower RMSEs than AE signals. This indicates that microphone and displacement signals may be more effective in predicting joint strength than AE signal. One possible reason is that the AE sensor was mounted on the anvil. The wave propagation through the specimens and anvil might introduce irrelevant information to the AE signals. On the other hand, the microphone and LVDT sensor are able to directly measure the vibrations of horn and specimens. In addition, AE sensors are known to be effective in micro-defects of solids [29]. Nevertheless, it is unclear if and

Table 3: Comparative results for the single-sensor case

Model	Features	Training Average (N)	Test Average (N)
SVR	Microphone	16.80	15.42
KNN	Microphone	14.94	16.62
LR	Microphone	16.32	15.96
SVR	Displacement	17.24	15.44
KNN	Displacement	17.12	16.12
LR	Displacement	16.88	16.04
SVR	AE	16.52	15.48
KNN	AE	15.30	17.14
LR	AE	16.28	16.08
ResNet20	Microphone	15.24	15.44
ResNet20	Displacement	15.60	15.68
ResNet20	AE	16.20	15.80

how micro-scale changes in materials influence the joint strength. This is worth further investigation.

4.3. Fusion strategy

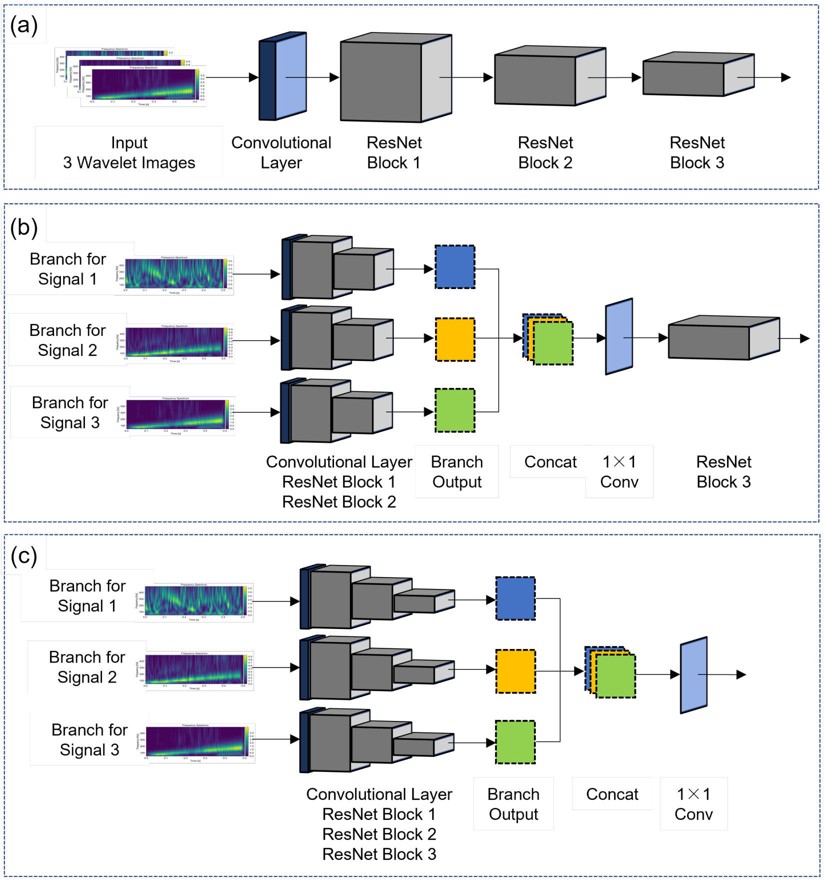

Fig. 10. Illustration of different sensor fusion strategies: (a) early fusion, (b) middle fusion, and (c) late fusion.

We explore three different fusion strategies, namely, early fusion, middle fusion, and late fusion, to fuse signals in the joint strength prediction model. The model structures for these fusion strategies are illustrated by Fig. 10. In the early fusion strategy, three signals are first stacked and then sent to the convolutional layer. Since the dimension of one wavelet transform image is 1×500×19, the dimension of the stacked input is 3×500×19. In the middle fusion strategy, three wavelet transform images go through different branches and the feature maps are extracted separately for different signals. The output feature maps of the second ResNet block are stacked together by a concatenation layer and sent to the third ResNet block for extracting feature maps after fusion. In the late fusion strategy, separate branches are used to extract feature maps from three different signals. Different from the middle fusion strategy, the fusion happens after the third ResNet block. After fusion, the combined feature map is sent to the fully connected layer to predict the joint strength. We use an additional convolutional layer with a 1×1 kernel to guarantee that the output channel matches with the previous network structure. It is worth noting that all three fusion strategies are flexible and can work with any number of signals.

Here, we explore all possible combinations of signals used for fusion. For each combination, we train and test three sensor fusion models (early, middle, and late). For conventional machine learning approaches, we perform feature selection for a feature pool that contains all features from all signals. The comparative results are summarized in Table 4.

The following observations can be drawn from Table 4. First, the best solution for UMW strength prediction is to fuse displacement and microphone signals using the late fusion strategy. It has the lowest average test RMSE (14.42 N) among all models. Second, deep learning models generally benefit from sensor fusion, because the test RMSEs generally reduce compared to single-sensor cases (see Table 3). However, conventional methods do not benefit from sensor fusion since their test RMSEs do not change much and even increase in some cases. This shows that deep learning model fuses data more effectively and is capable of extracting complementary information from different signals. Third, using fusion strategies properly is important. After applying fusion strategies, prediction models achieve lower test RMSEs in most cases. However, the late fusion strategy often has the lowest test RMSE among the three fusion strategies, which implies that fusing features that are processed by earlier network layers is beneficial for joint strength prediction. This could be attributed to the fact that the sensing signals used in the monitoring system are heterogeneous and may require different processing procedures. In the late fusion strategy, different sensing signals go through separate ResNet branches

Table 4: Comparative results for the sensor fusion case

Model	Features	Training Average (N)	Test Average (N)
SVR	All features	14.64	15.38
KNN	All features	12.90	16.54
LR	All features	14.50	17.06
Early fusion	Disp + Mic	15.76	15.36
Middle fusion	Disp + Mic	14.88	15.40
Late fusion	Disp + Mic	11.96	14.42
Early fusion	AE + Mic	14.28	15.92
Middle fusion	AE + Mic	14.28	15.82
Late fusion	AE + Mic	12.94	15.32
Early fusion	Disp + AE	12.58	15.10
Middle fusion	Disp + AE	13.36	15.60
Late fusion	Disp + AE	15.10	15.58
Early fusion	All signals	10.70	14.94
Middle fusion	All signals	13.76	15.72
Late fusion	All signals	12.22	14.78

(see Fig. 10), thus permitting different treatments of the signals. On the other hand, the early fusion strategy stacks wavelet images from three signals as inputs in the beginning and then the stacked images go through the same network layers including one convolutional layer and three ResNet blocks. Such processing cannot recognize the differences of different signals and may not be optimal. The middle fusion strategy uses separate branches to process wavelet images up to the second ResNet block and then fuses the intermediate outputs before the third ResNet block. This type of processing does not fully recognize differences between signal types either.

Fig. 11 shows the scatter plots of predicted vs. true joint strengths for the worst-performing and best-performing models. For the best-performing model, all predictions are close to the ground truth, indicating good prediction accuracy. The scatter plot of the worst-performing model depicts a random pattern instead of a linear trend. Some predictions deviate greatly from the true values. The worst model predicts more low-strength joints (e.g., <90 N) as high-strength (e.g., >110 N) than the best model. Likewise, the worst model also predicts more high-strength joints as low-strength than the best model.

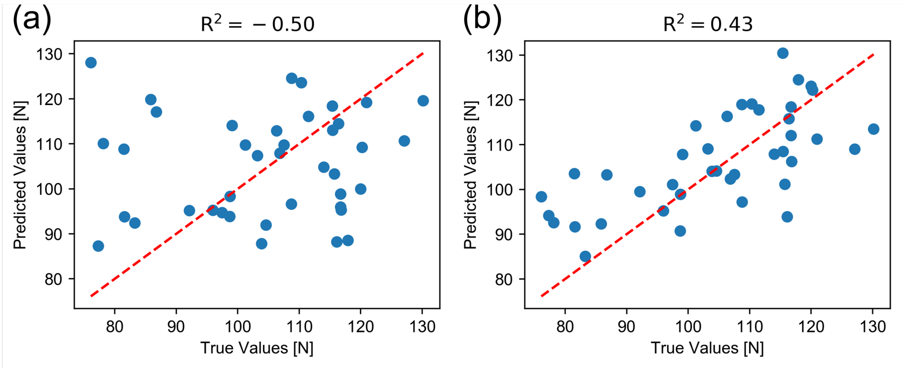

Fig. 11. Scatter plots showing predicted vs. true values in one test procedure for (a) the worst-performing model, i.e., LR with AE features ($R^2 = -0.50$) and (b) the best-performing model, i.e., the late fusion model using LVDT and microphone signals as inputs ($R^2 = 0.43$).

4.4. Correlation analysis of ResNet-generated and manual features

While deep learning has proven to be successful in a variety of fields, its interpretability and explainability still need investigation [30], especially in scientific, engineering, and manufacturing applications [31]. Since there is a lack of systematic methods for interpretable and explainable deep learning, we attempt to build a connection between RetNet-generated features and features extracted using physical knowledge with correlation analysis. In our ResNet20 model (see Fig. 3), the fully connected layer before "prediction" can be viewed as a set of 64 features. We calculate the Pearson's correlation coefficient between each pair of manual and ResNet-generated features. The results of the correlation analysis are reported by Figs. 12 and 13.

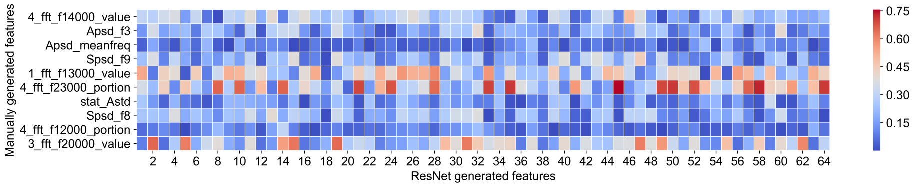

Fig. 12. A heatmap showing the correlations between the top 10 manual features selected by the conventional feature engineering method and 64 features automatically generated by ResNet.

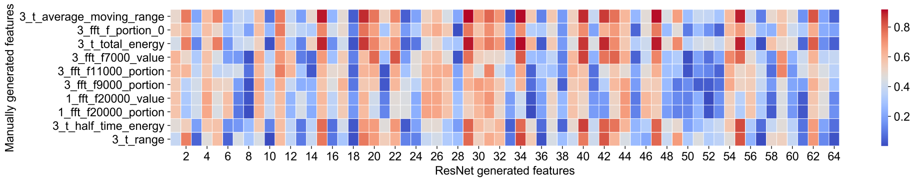

Fig. 13. A heatmap showing the correlations between the top 10 manual features with strongest correlations with 64 ResNet-generated features and corresponding correlations.

Fig. 12 displays a correlation heatmap for top 10 selected features by conventional machine learning methods (see Appendix for the implementation details). These 10 features are extracted using physical knowledge and selected according to their importance. As seen from Fig. 12, most of these features are highly correlated with at least one ResNet feature. Some features such as “1_fft_f13000_value,” “4_fft_f23000_portion,” and “3_fft_f20000_value” are strongly correlated with many ResNet features. This indicates that despite performing an end-to-end analysis of the input wavelet images, ResNet is able to resemble most physics-based features in an automatic fashion. Moreover, like other deep learning models, our ResNet20 model integrates feature extraction, feature selection, and regression modeling in a single architecture. This design enables an improved global solution. On the contrary, due to the extremely large solution space, most conventional machine learning methods perform feature extraction, feature selection, and regression modeling in a sequential way, so it is very challenging to find the optimal combination of features and regression models.

Another interesting observation is that most of the top 10 selected features are extracted from fast Fourier transform (FFT) of sensing signals and reflect vibration patterns. This validates our rationale that vibration patterns impact the joint formation and determine joint quality. As such, adopting wavelet transform to process sensing signals is a reasonable approach. Interested readers are referred to [14] for detailed descriptions of the manual features.

Fig. 13 shows 10 manual features with strongest correlations with ResNet features. These manual features contain information of time, energy, and frequency, and most of them are strongly correlated with multiple ResNet features. It demonstrates that ResNet can extract critical information reflecting physical knowledge of the UMW process.

4.5. Model convergence and computational cost

Common concerns for deep learning models, including ResNet, are the convergence in model training and expensive computational cost. Fig. 14 shows examples of one training loss curve and the RMSE curve calculated on the test set. During the training process, the test RMSE tends to converge to approximately 11 N as the training loss drops. It shows our model do not suffer from overfitting issues.

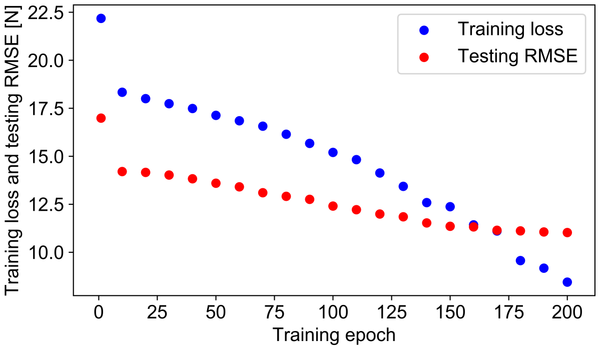

Fig. 14. Example training loss and test RMSE curves for ResNet training.

The deep learning models have different training time due to different model sizes. However, the training time for all deep learning models is below 10 minutes and the test time is below 1 second. The training time for other machine learning models is no more than 10 seconds and their test time is no more than 1 second. Despite being more computational expensive than conventional machine learning models, the proposed ResNet model can be conveniently trained with low computational resources, indicating its excellent accessibility.

5. Conclusion and future work

This paper presents an end-to-end online quality prediction method for UMW based on sensor fusion and deep learning. As shown by a case study using experimental data collected from an UMW process with four different tool conditions, the proposed approach outperforms conventional feature engineering and machine learning methods in terms of prediction accuracy. In addition, our method has the advantage of automatic feature generation/selection and offers an end-to-end solution for online UMW joint strength prediction. We present three lightweight sensor fusion strategies that can be conveniently incorporated into the prediction architecture. It is found that late fusion has the best prediction accuracy. We also observe that the proposed deep learning method is more effective than conventional feature engineering approaches in fusing data from multiple sensors.

Drawing on this work, several future research directions may be worth exploring. First, this study employs ResNet with 20 convolutional layers. The hyperparameters may be further tuned to improve the prediction performance. Some key parameters such as the number of convolutional layers, choices of activation layers, choices of loss function and learning rate can be carefully tuned to further optimize the network. Other neural networks may also be explored. Second, analyzing the feature maps generated by the ResNet in different welding stages may advance the physical understanding of the UMW process. Moreover, applying wavelet transform to the original signals instead of the truncated signals can preserve the high-frequency information contained in sensing signals, which may help improve the prediction performance. Third, as discussed in Section 4.4, interpretable and explainable deep learning is particularly important in manufacturing applications, but such methods are still lacking. As seen from the presented correlation analysis, ResNet is able to resemble important physics-based features. Future research efforts may be invested to reveal the underlying mechanism for how ResNet automatically generates features that preserve physical information. Physics-informed machine learning is also worth studying to enable the fusion of physical knowledge with data-driven approaches. Finally, though improvements are achieved compared to state-of-the-art methods, the prediction performance can be further improved by incorporating more variables that influence the joint quality (e.g., surface condition of specimens) as inputs. Such information will account for some variations in joint quality, thus helping improve the prediction accuracy.

6. Acknowledgments

This research has been supported by the National Science Foundation under Grant No. 1944345.

Appendix A. Feature extraction and selection in conventional machine learning methods

All the features developed in [14] are adopted by the conventional machine learning models in Section 4. In addition, we extract more features based on the observation reported in Section 3. It is seen that the energy for microphone and displacement signals is concentrated on frequencies that are multiples of 1000 between 5 kHz and 50 kHz. We calculate the total energy and corresponding energy proportion compared to the total energy in the signal every thousand from 5 kHz to 50 kHz. In total, more than 300 features are calculated from the sensing signals. However, not all all features are related to the joint quality and too many features may lead to overfitting. Most of the commonly used feature selection methods can be conveniently implemented using Python's sklearn library. Here, we use forward selection algorithm along with the mutual information calculation between features and joint strength residue. The feature selection procedure is briefly summarized as follows.

(1) All 200 groups of experimental data are randomly divided into training and test sets by a ratio of 4:1. Then we use the training set for feature selection.

(2) SVM with the RBF kernel is used as the regression model in the feature selection process. The feature subset is used as the input to train an SVM regression model. RMSE calculated using the current feature subset is used as the evaluation index.

(3) In each iteration of feature selection, the feature minimizing RMSE is selected as a new feature and added to the feature subset. The procedure is stopped when the RMSE is lower than a predetermined threshold.

References

[1] S. Shawn Lee, C. Shao, T. Hyung Kim, S. Jack Hu, E. Kannatey-Asibu, W. W. Cai, J. Patrick Spicer, J. A. Abell, Characterization of ultrasonic metal welding by correlating online sensor signals with weld attributes, Journal of Manufacturing Science and Engineering 136 (5).

[2] Z. Ni, F. Ye, Ultrasonic spot welding of aluminum alloys: A review, Journal of Manufacturing Processes 35 (2018) 580–594.

[3] G. Harman, J. Albers, The ultrasonic welding mechanism as applied to aluminum-and gold-wire bonding in microelectronics, IEEE Transactions on parts, hybrids, and packaging 13 (4) (1977) 406–412.

[4] I. Balz, E. Abi Raad, E. Rosenthal, R. Lohoff, A. Schiebahn, U. Reisgen, M. Vorländer, Process monitoring of ultrasonic metal welding of battery tabs using external sensor data, Journal of Advanced Joining Processes 1 (2020) 100005.

[5] T. Barnes, I. Pashby, Joining techniques for aluminium spaceframes used in automobiles: Part ii—adhesive bonding and mechanical fasteners, Journal of materials processing technology 99 (1-3) (2000) 72–79.

[6] L. Nong, C. Shao, T. H. Kim, S. J. Hu, Improving process robustness in ultrasonic metal welding of lithium-ion batteries, Journal of Manufacturing Systems 48 (2018) 45–54.

[7] Z. Ma, Y. Zhang, Characterization of multilayer ultrasonic welding based on the online monitoring of sonotrode displacement, Journal of Manufacturing Processes 54 (2020) 138–147.

[8] Y. Meng, C. Shao, Physics-informed ensemble learning for online joint strength prediction in ultrasonic metal welding, Mechanical Systems and Signal Processing 181 (2022) 109473.

[9] X. Shi, L. Li, S. Yu, L. Yun, Process monitoring in ultrasonic metal welding of lithium batteries by power signals, Journal of Manufacturing Science and Engineering 144 (5).

[10] Y. Meng, M. Rajagopal, G. Kuntumalla, R. Toro, H. Zhao, H. C. Chang, S. Sundar, S. Salapaka, N. Miljkovic, P. Ferreira, S. Sinha, C. Shao, Multi-objective optimization of peel and shear strengths in ultrasonic metal welding using machine learning-based response surface methodology, Mathematical Biosciences and Engineering 17 (6) (2020) 7411–7427.

[11] Y. Yang, C. Shao, Hybrid multi-task learning-based response surface modeling in manufacturing, Journal of Manufacturing Systems.

[12] C. Shao, W. Guo, T. H. Kim, J. J. Jin, S. J. Hu, J. P. Spicer, J. A. Abell, Characterization and monitoring of tool wear in ultrasonic metal welding, in: 9th international workshop on microfactories, 2014, pp. 161–169.

[13] C. Shao, T. H. Kim, S. J. Hu, J. J. Jin, J. A. Abell, J. P. Spicer, Tool wear monitoring for ultrasonic metal welding of lithium-ion batteries, Journal of Manufacturing Science and Engineering 138 (5).

[14] Q. Nazir, C. Shao, Online tool condition monitoring for ultrasonic metal welding via sensor fusion and machine learning, Journal of Manufacturing Processes 62 (2021) 806–816.

[15] Y. Zerehsaz, C. Shao, J. Jin, Tool wear monitoring in ultrasonic welding using high-order decomposition, Journal of Intelligent Manufacturing 30 (2) (2019) 657–669.

[16] L. Xi, M. Banu, S. Jack Hu, W. Cai, J. Abell, Performance prediction for ultrasonically welded dissimilar materials joints, Journal of Manufacturing Science and Engineering 139 (1).

[17] B. Zhou, M. Thouless, S. Ward, Predicting the failure of ultrasonic spot welds by pull-out from sheet metal, International journal of solids and structures 43 (25-26) (2006) 7482–7500.

[18] N. Shen, A. Samanta, W. W. Cai, T. Rinker, B. Carlson, H. Ding, 3d finite element model of dynamic material behaviors for multilayer ultrasonic metal welding, Journal of Manufacturing Processes 62 (2021) 302–312.

[19] T. H. Kim, J. Yum, S. J. Hu, J. Spicer, J. A. Abell, Process robustness of single lap ultrasonic welding of thin, dissimilar materials, CIRP annals 60 (1) (2011) 17–20.

[20] M. P. Satpathy, S. B. Mishra, S. K. Sahoo, Ultrasonic spot welding of aluminum-copper dissimilar metals: a study on joint strength by experimentation and machine learning techniques, Journal of Manufacturing Processes 33 (2018) 96–110.

[21] P. G. Mongan, E. P. Hinchy, N. P. O'Dowd, C. T. McCarthy, Quality prediction of ultrasonically welded joints using a hybrid machine learning model, Journal of Manufacturing Processes 71 (2021) 571–579.

[22] C. Shao, K. Paynabar, T. H. Kim, J. J. Jin, S. J. Hu, J. P. Spicer, H. Wang, J. A. Abell, Feature selection for manufacturing process monitoring using cross-validation, Journal of Manufacturing Systems 32 (4) (2013) 550–555.

[23] W. Guo, C. Shao, T. H. Kim, S. J. Hu, J. J. Jin, J. P. Spicer, H. Wang, Online process monitoring with near-zero misdetection for ultrasonic welding of lithium-ion batteries: An integration of univariate and multivariate methods, Journal of Manufacturing Systems 38 (2016) 141–150.

[24] E. B. Schwarz, F. Bleier, F. Guenter, R. Mikut, J. P. Bergmann, Improving process monitoring of ultrasonic metal welding using classical machine learning methods and process-informed time series evaluation, Journal of Manufacturing Processes 77 (2022) 54–62.

[25] M. J. Shensa, The discrete wavelet transform: wedding the a trous and mallat algorithms, IEEE Transactions on signal processing 40 (10) (1992) 2464–2482.

[26] S. Shawn Lee, T. Hyung Kim, S. Jack Hu, W. W. Cai, J. A. Abell, Analysis of weld formation in multilayer ultrasonic metal welding using high-speed images, Journal of Manufacturing Science and Engineering 137 (3).

[27] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[28] S. Targ, D. Almeida, K. Lyman, Resnet in resnet: Generalizing residual architectures, arXiv preprint arXiv:1603.08029.

[29] D. G. Aggelis, M. G. R. Sause, P. Packo, R. Pullin, S. Grigg, T. Kek, Y.-K. Lai, Acoustic Emission, Springer International Publishing, Cham, 2021, pp. 175–217.

[30] W. Samek, T. Wiegand, K.-R. Müller, Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models, arXiv preprint arXiv:1708.08296.

[31] R. Roscher, B. Bohn, M. F. Duarte, J. Garcke, Explainable machine learning for scientific insights and discoveries, Ieee Access 8 (2020) 42200–42216.