Physics-informed ensemble learning for online joint strength prediction in ultrasonic metal welding

Yuquan Meng[a], Chenhui Shao[a,*]

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

Abstract

Ultrasonic metal welding (UMW) is a promising solid-state joining technology that enables innovative and sustainable manufacturing. Despite possessing numerous advantages, UMW has a narrow operating window and is susceptible to both internal and external disturbances. As such, industrial scale UMW production calls for efficient, effective, and non-destructive joint quality assessment. To this end, this paper develops a novel hierarchical physics-informed ensemble learning (PIEL) framework that uses both physical knowledge and online sensing data for accurate online prediction of UMW joint strength. The PIEL framework decomposes the joint strength variability into a physics-informed global trend and a data-driven residual. The global trend is attributed to controllable or measurable welding conditions and is typically perceived as a large-scale variability. The data-driven residual is observed as a small-scale component when identical welding conditions are applied and can be captured by online sensing data. Drawing on this decomposition, hierarchical prediction models can be established to simultaneously account for both types of variabilities. As an essential component of the PIEL methodology, a highly efficient feature extraction procedure is developed using discrete wavelet transformation (DWT). The DWT-based feature extraction procedure is able to automatically extract key low-dimensional information from high-dimensional sensing signals, thus of-

*Corresponding author
Email address: chshao@illinois.edu (Chenhui Shao)

fering improved efficiency and effectiveness compared to conventional feature engineering approaches. Two real-world case studies with distinct physical setups are presented to demonstrate the effectiveness of the PIEL framework. The first case study investigates the influence of tool degradation and uses a dataset consisting of 200 welding cycles generated under four tool conditions. In the second case study, welding parameters including time, amplitude, and pressure are varied to generate 240 welds. When compared against multiple state-of-the-art baseline methods, PIEL consistently achieves superior prediction accuracy. Further, it is shown that by integrating physical knowledge, PIEL can effectively avoid overfitting and achieve excellent modeling robustness and data efficiency. While developed in the context of UMW, the PIEL framework is readily extensible to various other manufacturing processes.

Keywords: Physics-informed machine learning, Ensemble learning, Ultrasonic metal welding, Quality control

1. Introduction

Ultrasonic metal welding (UMW) is a solid-state joining technology identified as a key enabling technology to innovative and sustainable manufacturing [1]. UMW has wide industrial applications including automotive body construction [2, 3, 4], electric vehicle battery assembly [5, 6, 7], electronic packaging [8, 9, 10], and the assembly of hybrid heat exchangers [11]. Fig. 1 shows the configuration of a typical UMW system, which consists of generator, transducer, booster, horn, and anvil. In a welding cycle, UMW utilizes high-frequency ultrasonic vibrations to generate oscillating shears, which consequently joins the metal sheets clamped under pressure [12]. Compared to conventional fusion welding techniques, UMW provides various benefits such as energy efficiency, the ability to join dissimilar materials, environmental friendliness, and short welding cycles [1, 13].

Notwithstanding having numerous advantages, UMW has a narrow operating window and is known to be susceptible to both internal and external

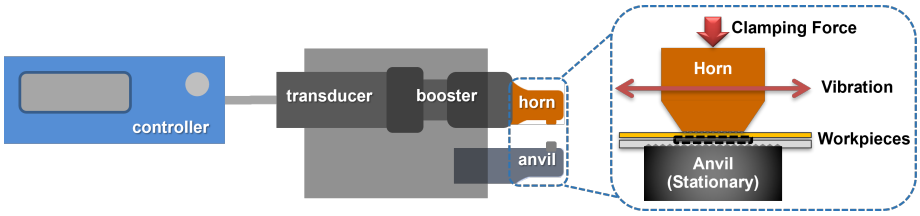


Figure 1: Configuration of a typical UMW system [14].

uncontrollable factors [7]. For example, the process physics of UMW is primarily determined by controllable welding parameters (e.g., welding time or welding energy, pressure, amplitude) and tool conditions [5, 14, 15, 16, 17, 18, 19, 20]. A relatively small set of parameter combinations can yield satisfactory joint quality for a certain application. Moreover, UMW is influenced by internal and external disturbances (e.g., specimen surface contamination [7, 21, 22]), which lead to substantial variabilities in joint quality. These uncontrollable factors are widely seen in real-world industrial applications but very challenging to monitor on the factory floor. As such, monitoring and predicting the quality of UMW joint are of vital importance in order to ensure the consistent joint quality in at-scale industrial productions [21, 22].

Existing works in quality prediction and control of UMW generally rely on (1) physics-based finite element (FE) simulations , e.g., [23, 24, 25] or (2) data-driven methods. FE simulations can provide some physical insights into the welding mechanism but are incapable of online, responsive prediction or adequately accounting for quality variabilities caused by uncontrollable factors [14]. Therefore, data-driven methods have attracted increasing attention in recent years. These methods have been mainly applied to response surface modeling [5, 14, 26, 27] and online process monitoring [16, 28, 22, 29, 30, 31].

Kim et al. [5] used response surface methodology to investigate the influence of welding time and pressure on the peel strength in UMW. Zhao et al. [26] deployed neural networks to construct the response surface for the joint strength of magnesium-titanium dissimilar welding. In [14], machine learning models are adopted to study the impact of amplitude and weld time on both peel and shear

strengths. Most recently, a hybrid multi-task learning model was developed in [27] to cost-effectively characterize UMW response surfaces through capturing the similarities between different material combinations.

Online process monitoring aims to detect process anomalies using in-situ sensing signals [28]. The first type of online monitoring methods targets on classifying joint quality into pre-determined classes (e.g., good vs. problematic) [21, 22, 28, 22, 29, 32]. For instance, the relationship between the weld attributes and two online signals including weld power and horn displacement was examined in [21]. Guo et al. developed an SPC-M online monitoring algorithm to detect the bad weld and their algorithm achieved almost zero misdetection rate [22]. Additionally, some studies use in-situ sensing data to predict tool conditions [16]. For example, Nazir and Shao developed an online sensing system and applied machine learning to successfully predict the tool conditions in a real-time fashion [16].

Despite the advancements brought by the abovementioned works, critical research gaps still exist. First, although response surface modeling can establish a connection between welding parameters and joint strength, it fails to capture the intra-variance of each welding parameter combination. This may be a serious drawback especially in factory-floor applications, where significant quality variabilities exist due to uncontrollable factors. Second, state-of-the-art online process monitoring approaches are restricted to classification tasks but fail to quantitatively predict the joint strength. The lack of detailed quality assessment prohibits more effective quality control for UMW.

To overcome these research gaps, this paper develops a hierarchical physics-informed ensemble learning (PIEL) framework for online joint strength prediction in UMW. The variability in joint strength is decomposed into a large-scale term characterizing the influence of physical conditions, such as tool degradation and welding parameters, and a small-scale residual term that is caused by uncontrollable factors and can be characterized by online sensing data. In addition, we develop an automatic feature extraction method using discrete wavelet transformation (DWT) that is applicable to all types of UMW sensing signals.

This method offers significantly improved efficiency compared to traditional feature engineering methods. Two case studies using different physical setups and datasets are presented to demonstrate the effectiveness of the proposed framework.

This remainder of this paper is organized as follows. Section 2 presents the details of the hierarchical PIEL framework and the modeling methods. Two case studies are discussed in Section 3. Finally, Section 4 concludes the paper and suggests future research directions.

2. Methodology

This section introduces details of the proposed hierarchical PIEL modeling framework for online joint strength prediction in UMW. Specifically, Section 2.1 presents the modeling framework. A highly efficient DWT-based feature extraction procedure is introduced in Section 2.2. Section 2.3 discusses the ensemble learning strategy and base models used in this study.

2.1. Hierarchical PIEL Framework

The quality variability in UMW can be attributed to two types of factors: (1) controllable or measurable process conditions and (2) uncontrollable disturbance factors. As shown by [14], controllable welding parameters lead to a large-scale variability, and when the same welding parameters are applied, small-scale variability is observed. As such, we propose a hierarchical modeling framework for UMW joint strength that is shown by Eq. (1) and illustrated by Fig. 2.

$$f(r, \mathbf{p}) = \mu(\mathbf{p}) + \eta(r; \mathbf{p}), \tag{1}$$

where $f(\cdot)$ is a function of the joint strength; $\mathbf{p}$ denotes the controllable or measurable process conditions; r is randomness caused by internal or external disturbances which are hard to measure; $\mu(\mathbf{p})$ is a global trend determined exclusively by $\mathbf{p}$; and $\eta(r; \mathbf{p})$ is the residual that relies on both $\mathbf{p}$ and r. The model of Eq. (1) decomposes the joint strength into two terms $\mu(\mathbf{p})$ and $\eta(r; \mathbf{p})$, which are discussed in more detail below.

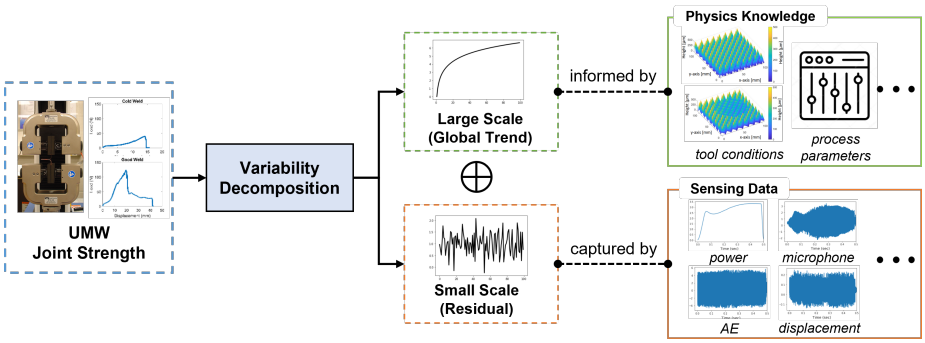


Figure 2: The hierarchical PIEL framework.

μ is perceived as a *global trend*. It is fully determined by controllable or measurable welding conditions that can be informed by physical knowledge on the UMW process. Examples include controllable welding parameters (e.g., welding time or welding energy, pressure, amplitude) and measurable tool conditions. It is worth noting that depending on the type of factors $\mathbf{p}$, different models should be used to model $\mu(\mathbf{p})$. For instance, if $\mathbf{p}$ is a set of welding parameters, $\mu(\mathbf{p})$ can be estimated using response surface methodology and regression models [5, 26, 14, 27]. If $\mathbf{p}$ denotes categorical tool conditions (e.g., new vs. worn), $\mu(\mathbf{p})$ is then a categorical term and can be characterized using a classification model [16, 33, 34].

η is a *residual* term representing the small-scale variability. This is similar with the natural or inherent variability in the philosophy of statistical process control. In other words, η denotes the unexplained variability when all controllable/measurable conditions are kept as constant. This type of variability can be captured by online sensing signals. A regression model is needed to map the relationship between sensing signals and η. In some cases, this relationship may be governed by process conditions $\mathbf{p}$ and η should be modeled as a function of both $\mathbf{p}$ and r, i.e., $\eta(r;\mathbf{p})$. Note that $\mathbf{p}$ serves as the hyperparameters of η and depending on value of $\mathbf{p}$, η may have a different form.

2.2. Automatic Feature Extraction Using DWT

This section develops an automatic feature extraction procedure using DWT. Three critical challenges exist in effectively and efficiently mining the online sensing signals in UMW. First, UMW experiments are costly and time-consuming, thus posing a high requirement on data efficiency [27]. The limited data size makes it extremely challenging to avoid overfitting while maintaining good performance. This also prohibits the application of some advanced machine learning methods such as deep learning. Second, the dimensionality of the as-received sensing signals is extremely high. Because UMW involves high-frequency vibration, a high sampling frequency must be used to capture the vibrational information. For example, in the case studies reported in Section 3, a sampling frequency of 250 kHZ was used in UMW experiments. Therefore, a 0.6 s welding cycle would have 150,000 data points in one single sensing signal. Third, due to the differences in the lengths of sensing signals, which are caused by different welding times, it is difficult to extract a set of features that are universally applicable to all welding cycles. For example, Fig. 3 shows three microphone signals obtained from UMW experiments with different welding times. The dimensionality varies significantly among the three signals.

To address these challenges, we use DWT to facilitate automatic feature extraction. While deep learning methods are able to provide an end-to-end solution and automatically accomplish feature extraction, feature selection, and model training, they typically require a sufficiently large training dataset, which is not available in many manufacturing applications. Here, we propose to automatically extract features using DWT for the following reasons. First, UMW has a high-frequency vibration nature and extracting frequency-domain features has been a common practice in the UMW quality monitoring literature. For example, frequency-domain analysis using fast Fourier transform was used to extract monitoring features from acoustic emission (AE) and microphone signals for UMW quality classification and tool condition monitoring [28, 22, 16]. Second, in existing works, time-domain features are mainly extracted from power and linear variable differential transformer (LVDT) signals. Time-domain features

that are proven to be effective, e.g., peak power and energy in the power signal, pre- and post-height in the LVDT signal [28, 21, 22, 16], are reflected by and can be reconstructed using some DWT features, so the key time-domain information is preserved in the DWT feature pool.

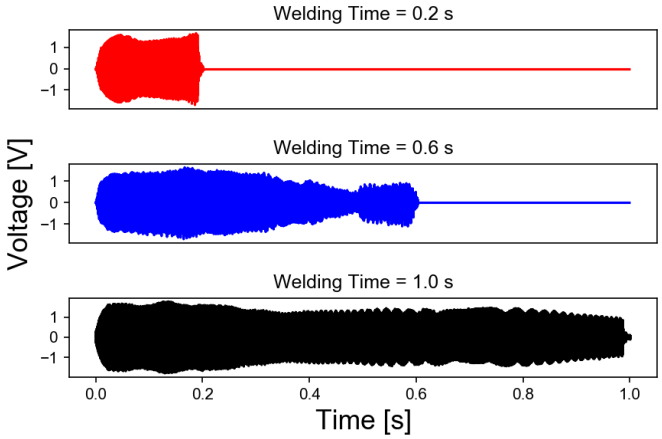


Figure 3: Comparison of microphone signals of different welding times.

The wavelet transform (WT) [35] can be described as

$$X_w(\tau, s) = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} x(t) \Psi \left(\frac{t-\tau}{s} \right) dt, \tag{2}$$

where τ is the shift parameter and s is the scale parameter; and $\Psi(t)$ is the mother wavelet which centers at zero with time support on a specific interval.

One common strategy of using WT is to obtain the spectrogram of the time series and then apply convolutional neural networks (CNN) or other neural networks on such spectrograms [36, 37, 38, 39]. However, deep neural networks working on images/spectrograms generally have complicated model architectures and require a large dataset for model training and hyperparameter tuning. As such, this type of strategy is less effective for data-scarce UMW applications. Here, we use DWT as a filter bank to decompose the signals into detail coefficients from the high pass filter and approximation coefficients from a low pass

filter. To do so, Eq. (2) is discretized as:

$$X_{m_d,n_d} = \int_{-\infty}^{+\infty} x(t)\Psi_{m_d,n_d}(t)dt, \tag{3}$$

$$\Psi_{m_d,n_d}(t) = \frac{1}{\sqrt{2^{m_d}}}\Psi\left(\frac{t - n_d 2^{m_d}}{2^{m_d}}\right), \tag{4}$$

where m_d is the level of decomposition, and n_d is the discretized shift parameter. In each level, the given time series with the highest frequency are decomposed into two branches. The resulting coefficients corresponding to high frequency are called detail coefficients (DC), whereas the remaining ones corresponding to low frequency are named as approximation coefficients (AC). This process can be repeated multiple times, until it reaches the specified decomposition level M_d. The most popularly adopted algorithm for DWT is Mallat algorithm [40], where the cutoff frequency goes half in levels and only approximation coefficients are further decomposed.

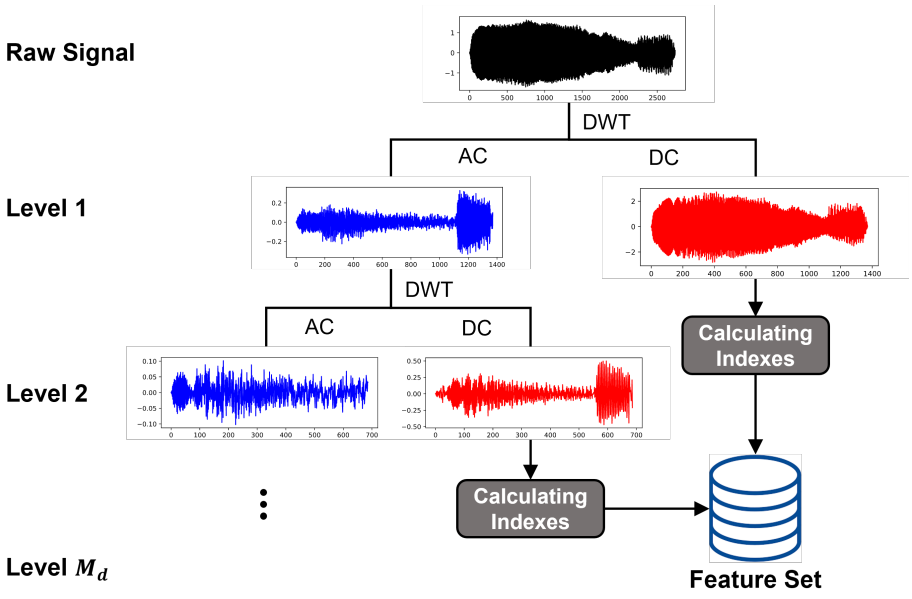


Figure 4: Illustration of implementing DWT on a microphone signal for feature extraction.

As illustrated by Fig. 4, a raw sensing signal is decomposed into several detail coefficients in levels, based on which the features are extracted. A list

of features extracted from each sensing signal is provided by Table 1. We note that extracting the DWT features does not require prior physical knowledge on the UMW process but can still effectively preserve both time-domain and frequency-domain information. Furthermore, the decomposition shown by Fig. 4 is not affected by the length of the raw signal.

Table 1: Definition of the DWT indexes.

Index	Definition
Entropy	the entropy of the sub-band
Variance	the variance of the sub-band
Mean	the mean of the sub-band
Median	the median of the sub-band
25th Percentile Value	the 25th percentile of the sub-band
75th Percentile Value	the 75th percentile of the sub-band
RMS	root mean square of the sub-band
Zero Crossing Rate	the number of times the signal crosses zero
Mean Crossing Rate	the number of times the signal crosses average

In UMW, there are mainly four online sensing signals, i.e., power signal, AE signal, microphone signal, and LVDT signal [16, 28]. The sensing signals collected by each sensor are processed individually by DWT and the indexes in levels are calculated respectively. All the indexes are then concatenated to form a feature pool. Though the dimensionality is significantly reduced compared to the raw signals, it is still too large to be directly used by a machine learning model. The total number of features is 672 when $M_d = 13$ (case study 1) and 624 when $M_d = 12$ (case study 2). Therefore, a feature selection procedure is required to further reduce the size of the feature set. Feature selection methods, typically categorized as wrapper methods, filter methods, and embedded methods [28, 41], should be selected depending on the scenarios. The specific feature selection procedure used in this work will be introduced in Section 3. After the feature selection, the resulting data will be fed into the ensemble learning

models. The whole procedure is illustrated by Fig. 5.

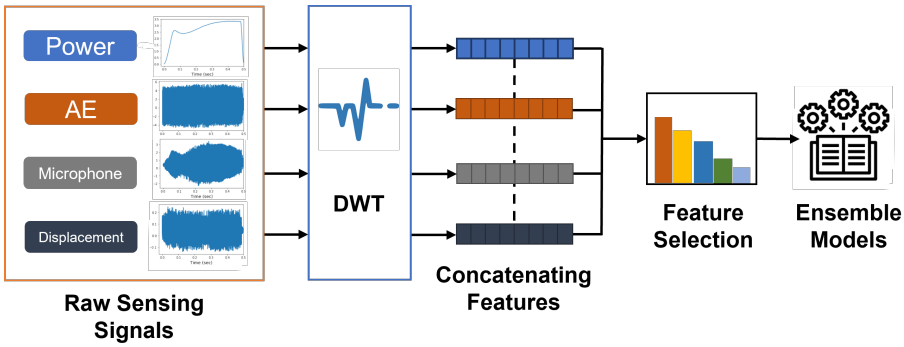


Figure 5: DWT-based feature extraction procedure.

2.3. Ensemble Learning

In our framework, ensemble learning along with cross validation (CV) is used to obtain good prediction accuracy while minimizing overfitting issues. Ensemble learning has been proven to be one of the most powerful machine learning methods, which dominates many Kaggle competitions. For example, Erickson et al. showed that the ensemble learning-based AutoML framework outperforms any other parameter tuning AutoML frameworks in most benchmark datasets [42]. The fundamental idea of ensemble learning is to train multiple base models and assemble them to form a more effective model [43, 44, 45].

The base models employed in this study are gradient boosting machines (GBM), which combines the predictions from multiple weak learners to generate the final results. The first explicit GBM algorithm was developed by Friedma [46]. GBM firsts train many weak learners, each of which may perform poorly on the task. Let $w_k(x;\theta_k)$ be the weak learner family where θ_k is the parameter set to be learned. Then given a training set $\{(x_i, y_i)\}_{i=1}^{n}$, a differentiable loss function $L(\cdot,\cdot)$ and number of iterations K, the algorithm, where F_k is the output of k-th iteration, can be summarized as follows:

Let

$$F_0 = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, \gamma), \tag{5}$$

where γ is a constant number.

Assuming that F_{k-1} is already obtained, define

$$\bar{y_i} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{k-1}(x)}, i = 1, 2, ..., n. \tag{6}$$

The parameter set for a weak learner $w_k(x; \theta_k)$ is evaluated by:

$$\theta_k = \arg\min_{\theta} \sum_{i=1}^{n} L(\bar{y_i}, w_k(x_i; \theta)). \tag{7}$$

The weight coefficient is obtained by:

$$\alpha_k = \arg\min_{\alpha} \sum_{i=1}^{n} L(y_i, F_{k-1}(x_i) + \alpha w_k(x_i; \theta_k)). \tag{8}$$

F_k is evaluated iteratively:

$$F_k(x) = F_{k-1}(x) + \alpha_k w_k(x_i; \theta_k). \tag{9}$$

Repeat this until $k = K$. Finally, the output of the model is F_K.

The philosophy of GBM is to train the weak learners iteratively such that new weak learner can compensate for the error led by all previous learners, through which the capability of the model is boosted. In UMW, the sensing signals have high dimensions, the training dataset is usually small, and the variability in joint strength is large, resulting in critical challenges for a single learning model to capture the relationship adequately. GBM can potentially overcome these challenges. There are many implementations of GBM, such as XGBoost and CatBoost. In this study, we select XGBoost, CatBoost, and light GBM with different sizes as the base model for the final ensemble models.

Fig. 6 shows the training process of ensemble models. Firstly, the dataset is divided into training and testing sets with the ratio of 7:3. Then k-fold CV is performed on the training set to train each base model, which is subsequently evaluated using the validation set. Based on the validation result, the hyper-parameters of base models are tuned to optimize the validation performance. The training-validation-tuning procedure is repeated until the validation performance is optimized.

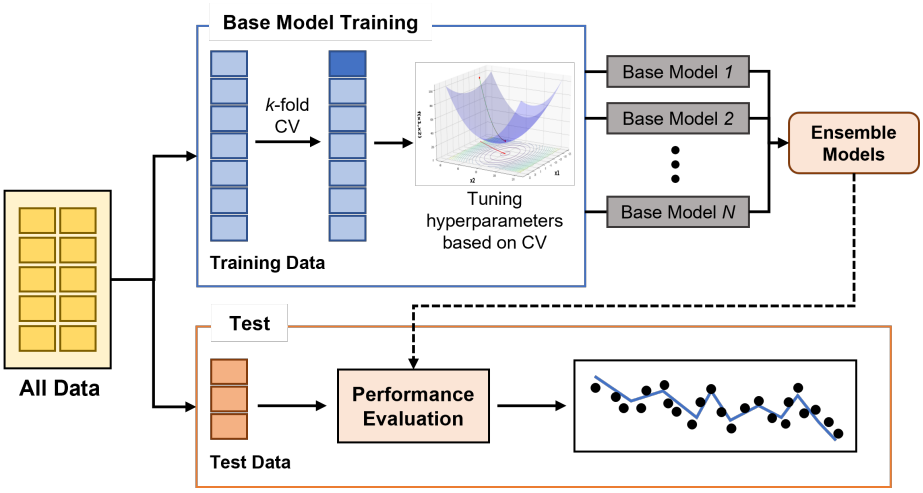


Figure 6: Training procedure for ensemble learning.

After each base model is trained separately, they are aggregated together to generate the final ensemble prediction model using the following weighted average aggregating rule:

$$M = \sum_{i=1}^{N} v_i M_i, \tag{10}$$

where M is the ensemble model, M_i is the ith base model, v_i is the weight assigned to base model M_i, $i = 1, \ldots, N$, and N is the number of base models. The weights v_i's are obtained according to each base model's learning performance using

$$v_i = \frac{\exp(-r_i)}{\sum_{i=1}^{N} \exp(-r_i)}, \tag{11}$$

where r_i is the validation RMSE for base model M_i. In other words, in the training and validation stage, base model M_i receives a score r_i indicating its performance. In this research, we use the RMSE value as a performance indicator, i.e., $r_i = \text{RMSE}_i$. This weighted averaging strategy can effectively alleviate overfitting issues. Note that there are other ensemble techniques such as blending, as well as other aggregating methods such as majority voting and averaging [47].

3. Case Studies

In this section, two case studies with different physical setups are investigated to demonstrate the effectiveness of the proposed PIEL framework. The first case study uses a dataset consisting of 200 UMW weld cycles, for which tool conditions are varied while the welding parameters are the same. This case study mimics an industrial scenario where the welding parameters have been optimized but tool degradation occurs during production. The second case study aims to predict the UMW joint strength when welding parameters are changed. This scenario refers to a practical application where the optimal welding parameters need to be determined and online monitoring is required to account for small-scale variability.

In both case studies, four sensors, namely, power, LVDT, AE, and microphone signals, are employed to collect the online sensing signals. Fig. 7 is a schematic of the online monitoring system used to generate the datasets. Fig. 8 shows examples of four sensing signals. The sampling rate was set as 250 kHZ in signal acquisition and the data collection duration was set as 2 s. Depending on the welding time selected, the segment corresponding to the vibration process will be different. The welding specimens were made of 110-copper and had a dimension of 50.8 mm $\times$ 25.4 mm $\times$ 0.2032 mm. After welding, all the joints were subjected to T-peel test [14] to measure the maximum peel load, which is defined as the joint strength. To rigorously examine the performance of PIEL, both case studies use various machine learning models such as random forest, support vector machine (SVM), CNN, and AdaLR as baseline methods. All the models are run on an Intel i9-7960X CPU @ 2.80GHz.

3.1. Joint Strength Prediction with Different Tool Conditions

In this case study, the PIEL framework is applied to predict the strength of joints made using the same welding parameters but the tool conditions are different. Table 2 shows the detailed welding configurations.

Fig. 9 shows the distributions of joint strength under different tool conditions. It is clear that tool conditions significantly influence the statistical

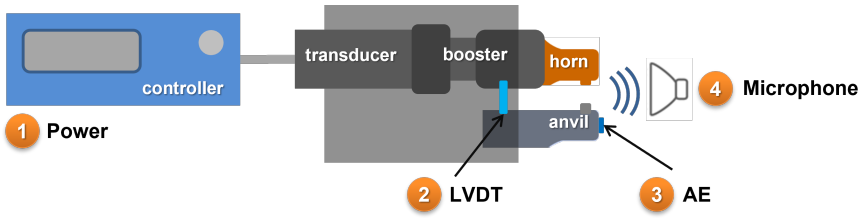


Figure 7: Schematic of the online UMW monitoring system. Adapted from [16].

Table 2: List of candidate machine learning models used in case study 1.

Tool Condition Label	Horn Condition	Anvil Condition	Welding Time	Welding Amplitude	Welding Pressure	Replicates
0	New	New	0.5	45	40	50
1	New	Worn				
2	Worn	New				
3	Worn	Worn				

distributions in terms of both mean and standard deviation. It is also worth noting that even under the same tool condition and welding parameters, the variability of joint strength is substantial, highlighting the necessity of online monitoring.

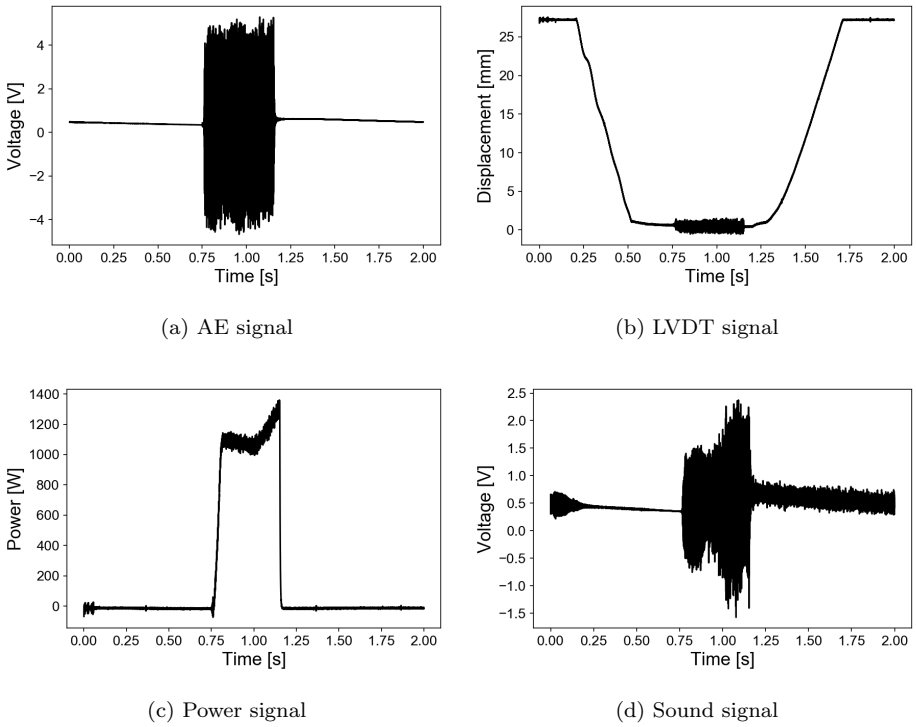
We adapt the hierarchical PIEL framework shown by Eq. (1) to obtain Eq. (12) for this case study.

$$f(r, p_{tool}) = \mu(p_{tool}) + \eta(r; p_{tool}), \tag{12}$$

where p_{tool} represents the tool condition and r is the disturbance. In real-world industrial applications, the tool condition is usually unknown. As such, it is desirable to establish a model which can predict the strength exclusively using online sensing signals. For this purpose, we obtain the following equation by assuming p_{tool} unknown

$$f(r, p_{tool}) = \eta(r; p_{tool}). \tag{13}$$

Since $\eta(r; p_{tool})$ is captured by online sensing signals, Eq. (13) raises the requirement that tool condition p_{tool} be inferred first by sensing signals and then



(a) AE signal

(b) LVDT signal

(c) Power signal

(d) Sound signal

Figure 8: Demonstration of typical sensing signals.

$\eta(r; p_{tool})$ can be obtained based on value of p_{tool}. To this end, the prediction process can be decomposed into two steps: (i) establishing a classification model to predict the tool condition based on sensing signals; and (ii) constructing ensemble models per each tool condition to predict the joint strength. Here, a multilayer perceptron (MLP) with features extracted by DWT as input is adopted for (i). We note that other classification models can be used for tool condition classification. In this study, we demonstrate an efficient way that does not require tedious feature engineering. The prediction procedure is illustrated by Fig. 10.

Following the training procedure of Fig. 6, 30% of data in each tool condition is selected as the test dataset. k-fold CV ($k = 5$), which divides the dataset as training and validation sets, is then performed on the remaining 70% data.

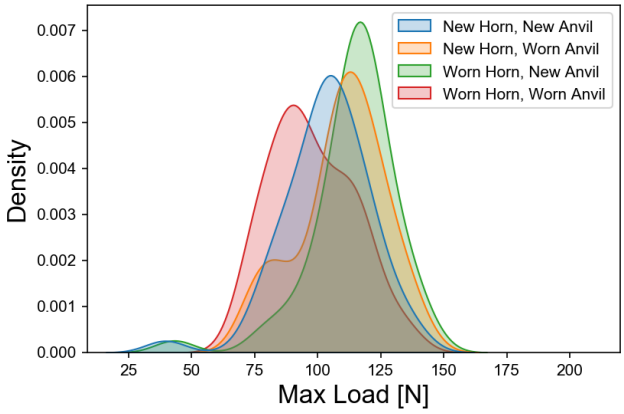


Figure 9: Kernel density plots for joint strength under different tool conditions.

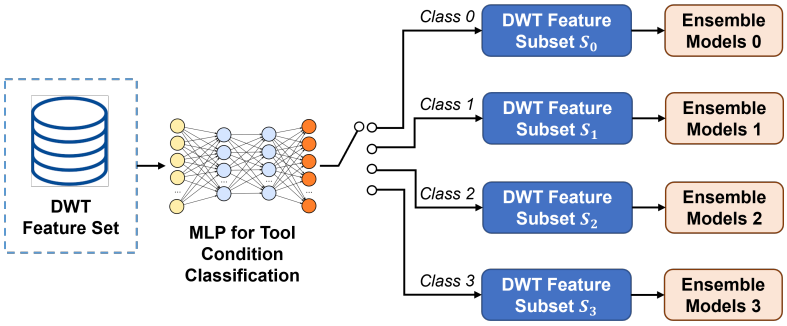


Figure 10: Prediction procedure for case study 1.

In total, 140 (35 per tool condition) and 60 (15 per tool condition) joints are included in the training and test sets. The 'rbio' wavelet family is selected as mother wavelet and 13 levels of decomposition are adopted for feature generation. The total number of candidate features is 672. The correlation is used for feature screening. Only the DWT features, whose absolute correlation coefficient with respect to the joint strength are greater than 0.35, are used by MLP and the ensemble models. The MLP used for tool condition classification has three fully connected layers with rectifier linear unit (ReLU) as the activation function and a batch normalization per-layer. The weak learners for base models are decision trees. The training algorithm follows the following procedure.

1. With DWT features, MLP is trained to predict the tool condition. Hyperparameters of MLP are tuned by validation scores.
2. Per each tool condition, feature selection is performed to obtain the DWT feature subset S_0, S_1, S_2, and S_3.
3. With input of S_i, the base models and ensemble models are trained for tool condition i (i = 0,1,2,3) independently, following the procedure of Fig. 6. Ensemble models for each tool condition are then obtained.

We construct baseline models by varying three conditions: (i) whether to classify tool conditions first, (ii) whether to use DWT features, and (iii) whether to use manual features. We denote $S = S_0 \cup S_1 \cup S_2 \cup S_3$, which is the union of feature sets selected for all four tool conditions. Manual features are extracted using conventional feature engineering methods. Two types of time-domain features are extracted from the power and LVDT signals. The first type is extracted using domain knowledge. Examples include welding energy and the change in horn displacement. The second type is generated using signal statistics such as mean, median, standard deviation, kurtosis, skewness, root mean square, maximum, and minimum. Because AE and microphone signals are periodic, frequency-domain features, such as peaks from the power spectral density, are mainly used for these two signals. A detailed description of the manual feature engineering process can be found in [16]. We denote S'_i as the manual feature subset obtained for tool condition i, $i = 0, 1, 2, 3$. $S' = S'_0 \cup S'_1 \cup S'_2 \cup S'_3$ is then the union of manual features selected for all tool conditions. Table 3 lists the size of feature subsets after feature selection.

A list of the constructed joint strength prediction models is provided by Table 4. PIEL-MF, PIEL-DWT, and PIEL-Comb use the same hyperparameters for ensemble models. Gradient-boosted tree is used as the weak learner of (1) XGBoost with a maximal depth of 6 and learning rate of 0.1, (2) lightGBM with learning rate of 0.05, and (3) CatBoost with a maximal depth of 6 and learning rate of 0.05. The number of weak learners is tuned by k-fold CV. XGBoost, lightGBM, and CatBoost models form the final ensemble prediction model fol-

Table 3: Size of the feature sets after feature selection.

Feature Set	Size	Feature Set	Size
S_0	51	S'_0	41
S_1	26	S'_1	27
S_2	26	S'_2	21
S_3	25	S'_3	23
S	117	S'	65

lowing Eqs. (10) and (11). GlobAvg is selected as a baseline model because online joint strength prediction is generally unavailable in most UMW systems and an average strength estimated from training phase is used to evaluate joint quality. When tool condition monitoring is available, one is able to obtain average joint strength per each tool condition, thus leading to ClassAvg, which is also selected as a baseline model.

Table 4: List of candidate machine learning models used in case study 1.

Model Name	Tool Condition Classification	Features Used	Description
PIEL-MF	Yes	S'	PIEL model using manual features only.
PIEL-DWT	Yes	S	PIEL model using DWT features only.
PIEL-Comb	Yes	$S \cup S'$	PIEL model using both manual and DWT features.
GlobAvg	No	None	The average strength of training set.
ClassAvg	Yes	None	The average strength in each tool condition of training set.
Random Forest	No	$S \cup S'$	Random forest with 100 estimators, minimal samples split of 2 and minimal samples leaf of 1.
SVM	No	$S \cup S'$	SVM with radial basis function (RBF) kernel.
AdaDT	No	$S \cup S'$	AdaBoost with 50 decision trees having a maximal depth of 3. Exponential loss function is used.

All the models are trained and tested on same training-testing split for five times. Benefiting from the efficient DWT-based feature extraction, all models can be trained in a fast manner—the training time of PIEL models is between 14 s and 20 s; and the training time of random forest, SVM, and AdaDT is

Table 5: Comparison of model performance for case study 1.

Model	Training RMSE	Test RMSE	Test R^2 Value
PIEL-MF	14.95	16.46	0.13
PIEL-DWT	13.95	15.84	0.21
PIEL-Comb	14.04	**15.40**	**0.24**
GlobAvg	NA	18.32	0
ClassAvg	NA	17.54	0.04
Random Forest	**5.92**	16.40	0.14
SVM	16.18	17.40	0.06
AdaDT	7.90	16.26	0.15

less than 1 s. Table 5 reports the comparative results. In classification, MLP with DWT features achieves 100% accuracy for training, validation, and test sets, demonstrating the effectiveness of DWT features in distinguishing tool conditions. PIEL-DWT and PIEL-Comb achieve the best prediction accuracy and significantly outperform the baseline models GlobAvg and ClassAvg. PIEL-Comb permits a 15.9% and 12.2% reduction in test RMSE compared to GlobAvg and ClassAvg, respectively. PIEL-DWT and PIEN-Comb have very close performance, indicating that automatic DWT-based feature extraction is highly effective and the generated DWT features are sufficient to ensure good performance. Additionally, all PIEL models have comparable training and test RMSEs, so they do not suffer from overfitting. On the contrary, random forest and AdaDT have severe overfitting issues. This demonstrates that by incorporating physical knowledge (tool condition classification in this case study), the model generalizability is greatly improved. One may notice that the R^2 values are small for all models. This is because the total sum of squares are small. While tool degradation leads to significant variability in joint strength, the variability is considerably smaller than that caused by changing process parameters, which is investigated in the second case study (Section 3.2). The small total sum of squares result in small R^2 values.

3.2. Joint Strength Prediction with Varied Process Parameters

The second case study focuses on a scenario where PIEL is used to simultaneously account for the influence of welding parameters and natural/inherent variability in UMW. 80 different welding parameter combinations were used to generate the UMW joints. Welding parameters and replicates are shown in Table 6. It should be noted that the signal length varies owing to different welding times. DWT-based feature extraction is used to accommodate this situation.

Table 6: UMW process parameters used in case study 2.

Parameter	Values
Welding Time	0.2 s, 0.4 s, 0.6 s, 0.8 s, 1.0 s
Welding Amplitude	30 µm, 35 µm, 40 µm, 50 µm
Welding Pressure	25 psi, 40 psi, 55 psi, 70 psi
Replicates	3

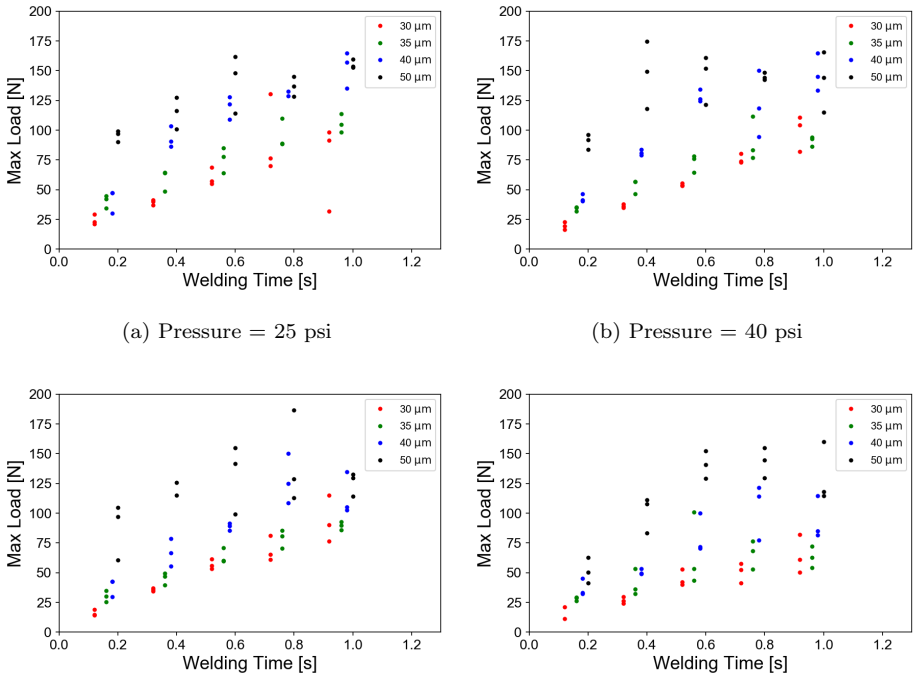
Fig. 11 visualizes the influence of welding parameters on the joint strength. It is seen that the joint strength has large intra-variance even when identical welding parameters are used, which again proves that it is imperative to account for such variabilities.

In this case study, Eq. (1) is modified to become

$$f(r, \mathbf{p}) = \mu(\mathbf{p}) + \eta(r), \tag{14}$$

where $\mathbf{p}$ represents the welding parameters, i.e., welding time, amplitude and pressure; and r represents the randomness that is captued by online sensing signals.

Similar to case study 1, 30% of total data is used as the test dataset. The training-validation-test framework depicted in Fig. 6 is used to train and tune the hyperparameters of ensemble models. Here, 'rbio' is used as mother wavelet and a DWT feature pool is established by decomposing data into 12 levels of approximation and detail coefficients. Feature importance [48] is employed to select the top 50 features from the feature pool. The ensemble learning models



(a) Pressure = 25 psi

(b) Pressure = 40 psi

(c) Pressure = 55 psi

(d) Pressure = 70 psi

Figure 11: Distributions of joint strength under different welding parameters.

use the same hyperparameters as those in case study 1. The global trend model and residual model are trained in sequence. The training algorithm is briefly summarized as follows.

1. An ensemble learning model with input $\mathbf{p}$ is first trained to estimate joint strength f, where k-fold CV is performed to tune the hyperparameters. The $\mu(\mathbf{p})$ model is obtained in this step.
2. Define residual as $\eta = f - \mu(\mathbf{p})$, and preliminary ensemble models predicting the residual with input of all DWT features are estimated using the training set.
3. By applying feature importance ranking, top 50 important features of the preliminary ensemble model are obtained.
4. With the feature set acquired from step 3, ensemble models are re-trained,

following the procedure of Fig. 6. The final $\eta(r)$ model is then obtained.

Table 7: List of candidate machine learning models used in case study 2.

Model Name	Description
PIEL	PIEL model using Eq. (14).
CNN	1-D CNN, including 2 convolutional layers with max pooling and 3 fully connected layers. ReLU is adopted as the activation function. Dropout and batch normalization are implemented per-layer.
Random Forest	Random forest with 100 estimators, minimal samples split of 2 and minimal samples leaf of 1.
AdaDT	Decision tree based AdaBoost. 50 decision trees with a maximal depth of 3 are selected as weak learners. Exponential loss function is used.
AdaLR	AdaBoost with 50 linear functions as weak learners. Squared loss function is used.
SVM	SVM with linear kernel and 0.9 as the regularization coefficient.

Table 8: Comparison of model performance in case study 2.

Model	Training RMSE	Test RMSE	Test R^2 Value
PIEL	13.19	**16.60**	**0.80**
CNN	23.25	28.06	0.41
Random Forest	6.96	23.59	0.68
AdaDT	5.55	23.07	0.61
AdaLR	**4.03**	23.97	0.42
SVM	15.75	19.18	0.73

370 The performance of the PIEL model is compared with state-of-the-art baseline models. As shown in Section 3.1, DWT features can preserve important information about quality. To investigate the effectiveness of the hierarchical PIEL model, we only use DWT features but not manual features in this case study.

Table 7 shows the definition of the models used in this case study. All models use both process parameters and DWT features as inputs, but only PIEL employs the hierarchical modeling structure given in Eq. (14). Five random training-test splits are generated to evaluate the performance of each model. Similarly with case study 1, the training of all models is computationally efficient—the training time for PIEL is ∼11 s; the training time for CNN is ∼1 min; and the training of random forest, AdaDT, AdaLR, and SVM takes less than 1 s. The training and test results are documented in Table 8. PIEL outperforms all the other models in terms of training RMSE, test RMSE, and test R^2 value. The poor performance of SVM reveals the curse of dimensionality, meaning that a model performs dramatically worse on high dimensional data (50 in this case). In contrast, PIEL shows excellent robustness to high dimensionality, further proving the advantage of integrating physical knowledge with ensemble learning.

4. Conclusion and Future Directors

Online joint strength prediction is critically needed in UMW production but has been an extremely challenging task. This paper presents a new hierarchical PIEL framework that enables efficient, accurate, and non-destructive online joint strength prediction. PIEL models the large-scale influence of physical factors and small-scale natural variability as a global trend and a residual, respectively. We also adopt a highly efficient feature generation method using DWT, which saves tedious feature engineering efforts and works with high-dimensional but small datasets. Two real-world case studies mimicking realistic production scenarios are reported. The first case study simulates a common production setting where welding parameters have been optimized but tool degradation is present and affects the joint quality. It is shown that PIEL can accurately predict both the tool conditions and the joint strength. PIEL achieves a 15.9% reduction in RMSE compared to the baseline method. The second case study investigates a complicated production setting where 80 different welding parameter combinations are used in UMW. PIEL outperforms all baseline methods and

permits a 13.5%–40.8% reduction in RMSE. Further, it is demonstrated that PIEL offers excellent generalizability, robustness, and data efficiency, owing to the integration of physical knowledge.

The developed hierarchical PIEL framework can be expanded to more complicated UMW production scenarios. In some production environments, there may exist more tool conditions induced by multiple levels or types of degradation in the horn and anvil [15]. Then we will need to extend the tool condition classification model such that it can recognize these tool conditions. Another interesting and important production scenario arises when both tool conditions and welding parameters are varying. In such cases, we can extend the PIEL model developed for case study 2 (Eq. (14)) by adding a hierarchy of tool condition monitoring. Then a hierarchical model can be used as the global trend model [49, 50, 51]. Despite that the PIEL framework is more data-efficient than state-of-the-art methods, the scalability may become a major challenge in these complicated scenarios. Hence, it is important to further improve the data efficiency and cost-effectiveness. Transfer learning of multiple production settings determined by different materials, tool conditions, etc. [52, 53, 54, 27, 55] may be a promising solution. We can also improve data efficiency by developing sampling design (aka active learning or intelligent design of experiment) algorithms to guide cost-effective data collection during the model building process [18, 20, 56].

CRediT Authorship Contribution Statement

Yuquan Meng: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. Chenhui Shao: Conceptualization, Methodology, Resources, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

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

References

[1] K. Martinsen, S. Hu, B. Carlson, Joining of dissimilar materials, Cirp Annals 64 (2) (2015) 679–699.

[2] A. Siddiq, E. Ghassemieh, Thermomechanical analyses of ultrasonic welding process using thermal and acoustic softening effects, Mechanics of Materials 40 (12) (2008) 982–1000.

[3] C. Zhang, D. Chen, A. Luo, Joining 5754 automotive aluminum alloy 2-mm-thick sheets using ultrasonic spot welding, Weld. J 93 (4) (2014) 131–138.

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

[5] 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.

[6] W. W. Cai, Ultrasonic Welding of Lithium-Ion Batteries, ASME Press, 2017.

[7] 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.

[8] H. Daniels, Ultrasonic welding, Ultrasonics 3 (4) (1965) 190–196.

[9] J. Kim, M. Chiao, L. Lin, Ultrasonic bonding of in/au and al/al for hermetic sealing of mems packaging, in: Technical Digest. MEMS 2002 IEEE International Conference. Fifteenth IEEE International Conference on Micro Electro Mechanical Systems (Cat. No. 02CH37266), IEEE, 2002, pp. 415–418.

[10] J. Kim, B. Jeong, M. Chiao, L. Lin, Ultrasonic bonding for mems sealing and packaging, IEEE Transactions on Advanced Packaging 32 (2) (2009) 461–467.

[11] G. Kuntumalla, Y. Meng, M. Rajagopal, R. Toro, H. Zhao, H. C. Chang, S. Sundar, S. Salapaka, N. Miljkovic, C. Shao, et al., Joining techniques for novel metal polymer hybrid heat exchangers, in: ASME International Mechanical Engineering Congress and Exposition, Vol. 59384, American Society of Mechanical Engineers, 2019, p. V02BT02A018.

[12] M. R. Rani, R. Rudramoorthy, Computational modeling and experimental studies of the dynamic performance of ultrasonic horn profiles used in plastic welding, Ultrasonics 53 (3) (2013) 763–772.

[13] S. Kumar, C. Wu, G. Padhy, W. Ding, Application of ultrasonic vibrations in welding and metal processing: a status review, Journal of manufacturing processes 26 (2017) 295–322.

[14] Y. Meng, M. Rajagopal, G. Kuntumalla, R. Toro, H. Zhao, H. C. Chang, S. Sundar, S. Salapaka, N. Miljkovic, P. Ferreira, et al., 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).

[15] C. Shao, T. H. Kim, S. J. Hu, 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).

[16] 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.

[17] 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.

[18] C. Shao, J. J. Jin, S. Jack Hu, Dynamic Sampling Design for Characterizing Spatiotemporal Processes in Manufacturing, Journal of Manufacturing Science and Engineering 139 (10), 101002. doi:10.1115/1.4036347. URL https://doi.org/10.1115/1.4036347

[19] 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.

[20] Y. Yang, Y. Zhang, Y. D. Cai, Q. Lu, S. Koric, C. Shao, Hierarchical measurement strategy for cost-effective interpolation of spatiotemporal data in manufacturing, Journal of Manufacturing Systems 53 (2019) 159–168.

[21] S. S. Lee, C. Shao, T. H. Kim, S. J. Hu, E. Kannatey-Asibu, W. W. Cai, J. P. 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).

[22] 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.

[23] 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).

[24] N. Shen, A. Samanta, H. Ding, W. W. Cai, Simulating microstructure evolution of battery tabs during ultrasonic welding, Journal of Manufacturing Processes 23 (2016) 306–314.

[25] D. Lee, E. Kannatey-Asibu Jr, W. Cai, Ultrasonic welding simulations of multiple, thin and dissimilar metals for battery joining, in: International Symposium on Flexible Automation, Vol. 45110, American Society of Mechanical Engineers, 2012, pp. 573–584.

[26] D. Zhao, D. Ren, K. Zhao, S. Pan, X. Guo, Effect of welding parameters on tensile strength of ultrasonic spot welded joints of aluminum to steel–by experimentation and artificial neural network, Journal of Manufacturing processes 30 (2017) 63–74.

[27] Y. Yang, C. Shao, Hybrid multi-task learning-based response surface modeling in manufacturing, Journal of Manufacturing Systems 59 (2021) 607–616.

[28] 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.

[29] W. Guo, J. Jin, S. Jack Hu, Profile monitoring and fault diagnosis via sensor fusion for ultrasonic welding, Journal of Manufacturing Science and Engineering 141 (8) (2019) 081001.

[30] 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.

[31] B. Wang, S. J. Hu, L. Sun, T. Freiheit, Intelligent welding system technologies: State-of-the-art review and perspectives, Journal of Manufacturing Systems 56 (2020) 373–391.

[32] B. Wang, Y. Li, Y. Luo, X. Li, T. Freiheit, Early event detection in a deep-learning driven quality prediction model for ultrasonic welding, Journal of Manufacturing Systems 60 (2021) 325–336.

[33] Y. Liu, L. Guo, H. Gao, Z. You, Y. Ye, B. Zhang, Machine vision based condition monitoring and fault diagnosis of machine tools using information from machined surface texture: A review, Mechanical Systems and Signal Processing 164 (2022) 108068.

[34] C. Zhou, K. Guo, J. Sun, Sound singularity analysis for milling tool condition monitoring towards sustainable manufacturing, Mechanical Systems and Signal Processing 157 (2021) 107738.

[35] J. Jin, J. Shi, Feature-preserving data compression of stamping tonnage information using wavelets, Technometrics 41 (4) (1999) 327–339.

[36] P. Wang, R. Yan, R. X. Gao, et al., Virtualization and deep recognition for system fault classification, Journal of Manufacturing Systems 44 (2017) 310–316.

[37] J. Wang, Y. Ma, L. Zhang, R. X. Gao, D. Wu, Deep learning for smart manufacturing: Methods and applications, Journal of manufacturing systems 48 (2018) 144–156.

[38] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R. X. Gao, Deep learning and its applications to machine health monitoring, Mechanical Systems and Signal Processing 115 (2019) 213–237.

[39] Y. Kim, K. Na, B. D. Youn, A health-adaptive time-scale representation (htsr) embedded convolutional neural network for gearbox fault diagnostics, Mechanical Systems and Signal Processing 167 (2022) 108575.

[40] S. G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, in: Fundamental Papers in Wavelet Theory, Princeton University Press, 2009, pp. 494–513.

[41] J. Tang, S. Alelyani, H. Liu, Feature selection for classification: A review, Data classification: Algorithms and applications (2014) 37.

[42] N. Erickson, J. Mueller, A. Shirkov, H. Zhang, P. Larroy, M. Li, A. Smola, Autogluon-tabular: Robust and accurate automl for structured data, arXiv preprint arXiv:2003.06505.

[43] D. Zhang, P. Baraldi, C. Cadet, N. Yousfi-Steiner, C. Bérenguer, E. Zio, An ensemble of models for integrating dependent sources of information for the prognosis of the remaining useful life of proton exchange membrane fuel cells, Mechanical Systems and Signal Processing 124 (2019) 479–501.

[44] F. Zhang, J. Yan, P. Fu, J. Wang, R. X. Gao, Ensemble sparse supervised model for bearing fault diagnosis in smart manufacturing, Robotics and Computer-Integrated Manufacturing 65 (2020) 101920.

[45] V. Yaghoubi, L. Cheng, W. Van Paepegem, M. Kersemans, An ensemble classifier for vibration-based quality monitoring, Mechanical Systems and Signal Processing 165 (2022) 108341.

[46] J. H. Friedman, Stochastic gradient boosting, Computational statistics & data analysis 38 (4) (2002) 367–378.

[47] O. Sagi, L. Rokach, Ensemble learning: A survey, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8 (4) (2018) e1249.

[48] M. Saarela, S. Jauhiainen, Comparison of feature importance measures as explanations for classification models, SN Applied Sciences 3 (2) (2021) 1–12.

[49] S. Raudenbush, A. S. Bryk, A hierarchical model for studying school effects, Sociology of education (1986) 1–17.

[50] A. Gelman, Multilevel (hierarchical) modeling: what it can and cannot do, Technometrics 48 (3) (2006) 432–435.

[51] Y. Yang, D. J. McGregor, S. Tawfick, W. P. King, C. Shao, Hierarchical data models improve the accuracy of feature level predictions for additively manufactured parts, Additive Manufacturing (2022) 102621.

[52] S. J. Pan, Q. Yang, A survey on transfer learning, IEEE Transactions on knowledge and data engineering 22 (10) (2009) 1345–1359.

[53] C. Shao, J. Ren, H. Wang, J. J. Jin, S. J. Hu, Improving machined surface shape prediction by integrating multi-task learning with cutting force variation modeling, Journal of Manufacturing Science and Engineering 139 (1).

[54] H. Chen, Y. Yang, C. Shao, Multi-task learning for data-efficient spatiotemporal modeling of tool surface progression in ultrasonic metal welding, Journal of Manufacturing Systems 58 (2021) 306–315.

[55] M. Ramezankhani, B. Crawford, A. Narayan, H. Voggenreiter, R. Seethaler, A. S. Milani, Making costly manufacturing smart with transfer learning under limited data: A case study on composites autoclave processing, Journal of Manufacturing Systems 59 (2021) 345–354.

[56] M. Mehta, C. Shao, Adaptive sampling design for multi-task learning of gaussian processes in manufacturing, Journal of Manufacturing Systems 61 (2021) 326–337.