Explainable Few-Shot Learning for Online Anomaly Detection in Ultrasonic Metal Welding with Varying Configurations

Yuquan Meng[a], Kuan-Chieh Lu[a], Zhiqiao Dong[a], Shichen Li[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 solid-state joining technology with widespread industrial applications. While UMW has numerous important advantages compared to traditional fusion-based welding methods, its performance can be substantially influenced by process anomalies such as tool degradation and material surface contamination, which are commonly encountered in industrial-scale productions. Recently, online monitoring has demonstrated excellent anomaly detection capabilities. However, the existing monitoring algorithms require a large amount of labeled data and lack the generalizability or adaptability to new process configurations (i.e., domains). This paper develops a meta-learning-based explainable few-shot learning (XFSL) framework that enables highly data-efficient adaptation of online monitoring algorithms to new process configurations with excellent explainability. We consider two distinct types of problems including tool condition monitoring and workpiece surface condition classification with varying UMW configurations. Using experimental data, we demonstrate that the proposed XFSL method achieves high classification performance in previously unseen target domains and significantly outperforms baseline methods. Furthermore, XFSL is able to evaluate the importance

*Corresponding author

Email addresses: yuquanm2@illinois.edu (Yuquan Meng), kclu3@illinois.edu (Kuan-Chieh Lu), zhiqiao5@illinois.edu (Zhiqiao Dong), shichen8@illinois.edu (Shichen Li), chshao@illinois.edu (Chenhui Shao)

of each feature, thus revealing key features, feature types, and signal frequencies. It is shown that explainability-based feature selection can effectively eliminate unimportant information from monitoring signals while maintaining and even improving prediction performance. The proposed XFSL method is extensible to other manufacturing applications and holds significant potential for advancing the generalizability, adaptability, and agility of decision-making algorithms in modern manufacturing.

Keywords: few-shot learning, meta-learning, domain adaptation, anomaly detection, ultrasonic metal welding, quality control, explainable machine learning

2010 MSC: 00-01, 99-00

---

1. Introduction

Ultrasonic metal welding (UMW) is a versatile solid-state welding technology that can join multiple sheets, wires, or foils of similar or dissimilar metals [1, 2]. UMW has numerous industrial applications, including automotive body construction [3–5], lithium-ion battery assembly [1, 6, 7], and electronic packaging [8]. Compared with conventional fusion welding methods, UMW offers important advantages including environmental friendliness, energy efficiency, and high production rate [1, 9, 10].

One of the critical challenges in the industrial implementation of UMW is the presence of strong process variability. UMW usually has small operating windows, especially when the workpieces are thin, dissimilar metals, such as in battery assembly applications [10, 11]. Moreover, UMW is sensitive to both internal and external process anomalies including but not limited to tool degradation [12–14] and surface contamination of workpieces [7, 15, 16]. Responsively detecting process anomalies is important for not only process monitoring but also real-time control [7, 16]. As such, online monitoring has attracted extensive attention over the last decade, e.g., [7, 10, 11, 14–19]. For example, online quality monitoring methods have been developed to classify joint quality [11, 17] or

predict joint strength [10, 19]. Nazir and Shao [14] developed an online tool condition monitoring (TCM) system based on sensor fusion and machine learning that achieved near-perfect classification accuracy. For workpiece surface condition monitoring, Lee et al. [15] determined the correlations between online sensing signals, process conditions, and joint performance. They also demonstrated the feasibility of detecting surface contamination online. Recently, Lu et al. [16] created an online cost-effective approach for classifying mixed tool and material conditions, and investigated the cost factors of online monitoring including sensors, signal fraction, and sampling rate.

Recent research in UMW process monitoring has shown the effectiveness of machine learning algorithms in accurate predictions [20]. These machine learning methods require a large amount of labeled training data collected from fixed process configurations (i.e., domains). However, modern manufacturing is featured by rapid reconfiguration and agile adaptation [21, 22] that necessitate varying process configurations. In UMW applications, a process configuration can be defined by materials, welding parameters, surface conditions, etc. Changing process configurations may lead to very different data distributions. In such cases, it is necessary to collect labeled data from the new scenarios and re-train classification algorithms or even build completely different algorithms. Acquiring a significant volume of training data can be time-, labor-, and resource-intensive, which is highly undesirable in manufacturing. As such, the ability to effectively transfer knowledge from rich-resource and established domains to data-scarce domains plays an important role in the smart manufacturing paradigm [23–26].

Few-shot learning (FSL) [27], which aims to train models that can perform well on new tasks with minimal data, is a promising solution to the aforementioned issues. Model-Agnostic Meta-Learning (MAML) is one of the most popular FSL methods. As a meta-learning approach, MAML is able to train machine learning models that are conveniently adaptable to new tasks. In addition, MAML offers a general framework that can be integrated with various model architectures and problem settings [28]. MAML trains a model on a vari-

ety of tasks and uses gradient descent to optimize the model's initial parameters such that the model can be adapted to a new task with only a few gradient updates. MAML has been shown to achieve strong results in a variety of FSL tasks, e.g., cross-lingual transfer [29] and semantic feature learning for domain generalization [30].

Most recently, FSL has started receiving some attention in manufacturing. For instance, Liu et al. [31] developed an MAML-based meta-invariant feature space method to estimate the tool wear in cutting processes under various data-deficient cutting conditions. Wang et al. [32] combined meta-learning with fine tuning to predict cutting tool wear. Nevertheless, research on FSL in manufacturing has been limited to machining processes, but no studies are available for UMW or other manufacturing processes. TCM for UMW is more challenging than machining because of the complicated process physics and limited understanding on the tool wear mechanisms [12, 14]. Moreover, the existence of other process anomalies such as workpiece surface contamination introduces additional challenges to online monitoring [7, 16].

One critical limitation of existing FSL methods is the lack of interpretability or explainability due to their inherent "black-box" nature. Such interpretability is of vital importance in manufacturing because it helps reduce the risk of errors and improve the reliability as well as safety of the manufacturing process. One potential solution to overcome this limitation of FSL is Explainable Artificial Intelligence (XAI). In general, XAI methods can be classified into three types: saliency maps, attention mechanisms, and rationale models [33]. Among those methods, saliency maps, which visualize the gradient of the model's output probability of a class with respect to its input, have drawn most attention. Studies such as Simonyan et al. [34] and Zeiler et al. [35] have shown the utility of saliency maps in identifying relevant patterns and objects in images, guiding the optimization of complex models. In manufacturing-related fields, Su et al. used saliency map to identify the machine area as an input of the downstream models [36]. Ye and Yu proposed a saliency map based convolutional neural network to detect machine fault [37]. However, two challenges must be addressed

when applying saliency map to FSL. First, the traditional saliency map methods cannot be directly integrated with FSL because they are local explanation methods and confined to visualizing feature importance for individual instance instead of a task. Moreover, the randomness of model training in FSL could lead to large variations of feature importance, prohibiting its applications as the feature ranks are undetermined.

To address the aforementioned research gaps, this paper develops an explainable few-shot learning (XFSL) framework for online anomaly detection in UMW with varying configurations. XFSL can effectively transfer the meta-knowledge learned from data-rich domains to data-scarce domains. Explainability is enabled by using the extracted meta-knowledge to evaluate the feature importance and extending saliency map approaches. The new explainability approach can capture the global characteristics of each class and is less sensitive to the randomness of model training. To evaluate the performance of the proposed framework, a series of case studies are conducted with distinct UMW process configurations and mixed anomalies. Results show that the proposed XFSL method achieves high classification accuracy and significantly outperforms baseline methods, indicating XFSL can serve as an effective feature selection method. The explainability-based feature selection effectively eliminates unimportant information from monitoring signals while maintaining and even improving prediction performance.

The remainder of the paper is organized as follows. Section 2 formulates the problem and presents the details of the XFSL framework. Section 3 reports the data acquisition process and case studies. Finally, Section 4 concludes the paper.

2. UMW Process and Online Monitoring System

2.1. UMW overview

Fig. 1 illustrates a typical UMW process. In a typical UMW cycle, the workpieces are clamped by welding tools, including the stationary anvil at the

bottom and the oscillating horn at the top. When the process starts, a pneumatic system generates a sufficient clamping pressure to act as the normal force. Then, the high-frequency vibrations of the horn, which are actuated by a transducer and a booster, remove surface oxides and initiate bonding between metal interfaces. Finally, the deformation layer starts growing, and a bond is formed by the compound effect of mechanic interlocking and atomic diffusion. During the process, vibration amplitude is maintained constant by a built-in controller, and instantaneous power consumption is recorded by an internal power meter. Ho ıs, an

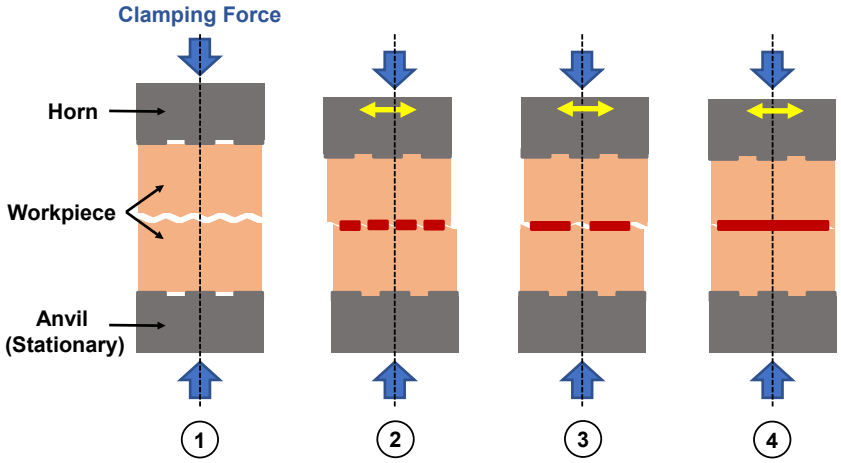

Fig. 1. Illustration of a typical UMW process [16].

The tools of UMW, i.e., horn and anvil, play a critical role in joint strength and welding quality [14, 38]. Both tools have many pyramid-shaped knurl patterns on the surface, and these patterns are the key elements to effectively transferring energy from the horn to workpieces for bonding formation. However, knurl patterns wear down gradually as more welds are produced. Shao et al. [12, 38] characterized the geometric changes of anvil knurls on anvil and showed that the tool wear progression in UMW is highly complicated. The changes in tool surface geometry may influence the vibrational patterns of metal

layers and therefore significantly affect joint quality [14].

2.2. Online monitoring system

Experiments in this research were conducted using a Branson Ultraweld L20 UMW machine, which is equipped with a customized online monitoring system, as depicted in Fig. 2. The data acquisition (DAQ) system has four online sensors, which are briefly introduced as follows.

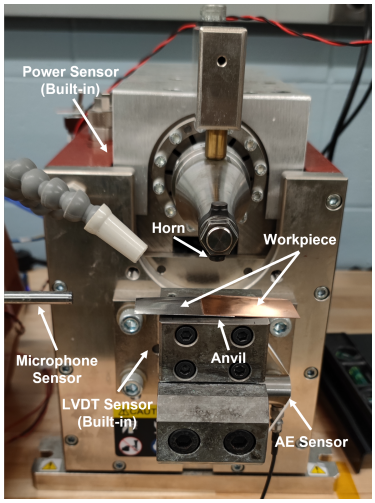

Fig. 2. Photo of the UMW machine with sensors.

1. Acoustic emission (AE): A Physical Acoustic-R15α AE sensor attached to the anvil assembling collects acoustic signal during the welding cycle. The signal is amplified by a Physical Acoustic 2/4/6 voltage preamplifier.

2. Linear variable differential transformer (LVDT): A built-in LVDT measures the vertical relative displacement between the horn and stationary anvil.

3. Power meter: A built-in power meter records the instantaneous power consumption of the welder during the welding cycle.

4. Microphone: A GRAS 40PP microphone mounted close to the welding area records sound signals during the welding cycle. The signal is amplified by GRAS 12AL preamplifier.

Monitoring of the process anomalies with online signals stands as a critical yet challenging task [13, 38, 39]. One important prerequisite to the success of online process monitoring is that the implemented sensors can sufficiently capture changes in process conditions [13, 39, 40]. In UMW, there are mainly two process anomalies: tool degradation [12–14] and workpiece surface contamination [7, 15, 16]. Some prior research has shown the feasibility of employing sensing signals for process monitoring and developed monitoring algorithms. For instance, Nazir and Shao [14] studied the signals collected and concluded the signals contain rich information about the process which can be used for tool condition monitoring. It was also shown that online sensing signals were able to reflect workpiece surface conditions [15, 16].

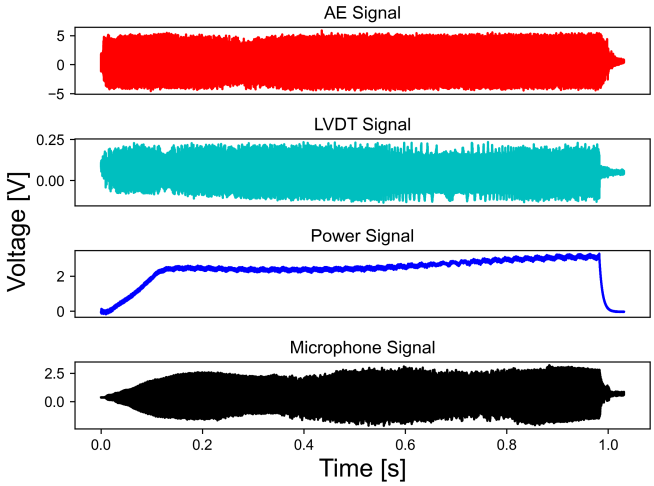

Fig. 3. Examples of raw online sensing signals.

However, state-of-the-art research typically requires a large amount of labeled training data, mainly because of the challenges encountered in processing signals. First, the collected sensing signals are extremely high-dimensional. To

sufficiently capture the high-frequency vibrational information, a sampling frequency of 250 kHz was used by the DAQ system in this study. Fig. 3 shows example raw signals from each sensor. With a 1 s welding time, the signal dimension is 250,000. Second, it is critical to capture the changes of the signal patterns over time, because those changes are related to the progressive formation of the joint and can characterize different stages of the process. Nevertheless, as shown by Fig. 3, the signal patterns are complicated and do not support a clear pathway for tracking the process progression. Third, four sensing signals used by the monitoring system are heterogeneous and their correlations are unclear, making their processing very challenging. As such, substantial training data is required for a machine learning model to sufficiently learn the intricate relationships between sensing signals and process conditions. Moreover, the complexity of the mapping from the sensing signals to the process physics introduces significant challenges to the development of monitoring algorithms in data-scarce scenarios.

In this work, the automatic feature extraction procedure developed in [10, 16] is employed to extract time-frequency features from online sensing signals. Specifically, discrete wavelet transform (DWT) is used to decompose the original signals into 13 levels of wavelet coefficients. For each sensor, 12 different statistical indexes are computed from each level of wavelet coefficients including entropy, zero crossing rate, mean crossing rate, 5th/25th/75th/95th percentile, median, mean, standard deviation, variance, and root mean square value. The complete explanation of each index can be found in [10]. Therefore, a total of 624 features are generated for each data instance and greatly reduce the dimensionality of the input feature space.

3. Overview of the Proposed Method

3.1. Problem formulation

In order to present the developed model for FSL in UMW, it is necessary to formulate the machine learning problem which accurately represents the physical process. The terminology is defined as follows.

- *Domain*: A domain is a joint distribution P_{XY} on $X \times Y$, where X is the input feature space and Y is the label space. For example, in a TCM problem, X is the vector of features extracted from online sensing signals. Y is the tool condition label. Likewise, Y is the label for workpiece surface conditions if the goal is to detect surface contamination.

- *Task*: Each task is an individual learning problem that meta-learning learns to tackle, where a specific classifier (inner-learner) is trained exclusively for fulfilling the task.

- *Support set/query set*: In each task, support set D_{sup} is used to train the inner-learner, similar to the concept of training data in conventional machine learning, whereas query set D_{query} is for evaluating task loss and updating meta-knowledge, i.e., initializations.

- *Meta-train/meta-validate/meta-test*: Learn/validate/test meta-knowledge on a series of tasks.

- *Inner loop/outer loop*: Inner loop is the phase when inner-learner is trained in each task. In contrast, outer loop consists of meta-train and meta-validation.

Suppose the set of source domains D_{source} consists of k different domains D^1, D^2, ..., D^k. Let D_{target} be the target domain. Therefore, domain adaption can be formulated as a supervised learning problem, where the goal is to learn a model that can generalize from D_{source} to D_{target} given a limited number of labeled examples in D_{target}. Note that the distributions of domains D^1, D^2, ..., D^k and D_{target} may be significantly different. For example, in the TCM task, the online sensing signals along with corresponding tool condition labels under the same process configuration constitute a domain. Whereas the signals generated under different process configurations belong to different domains. It is expected that the P_{XY} distribution is different for different domains. Therefore, the classification models trained for the source domains D_{source} are not directly transferable to the target domains D_{target}. On the other hand, despite

the differences, because these domains share the common process physics, i.e., UMW, they share some similarities, making it feasible to transfer knowledge between them.

3.2. Overview of XFSL

Fig. 4 shows a schematic overview of the XFSL framework. The framework mainly includes two components: the FSL method and the global explanation approach. FSL extracts and learns the meta-knowledge that is shared across source domains. The meta-knowledge represents the common patterns and inherent characteristics that are universally applicable across domains. Building on the meta-knowledge, FSL can effectively generalize the knowledge (models) to new, unseen scenarios. The global explanation method leverages the meta-knowledge extracted by FSL to attribute the input features. By utilizing the learned meta-knowledge on the data distribution of the target domain, the global explanation method estimates the global contribution of each input feature, i.e., feature importance. The availability of feature importance scores facilitates the identification of crucial features, feature types, and signal frequencies as well as an effective feature selection procedure. The technical details of XFSL will be presented in Section 4.

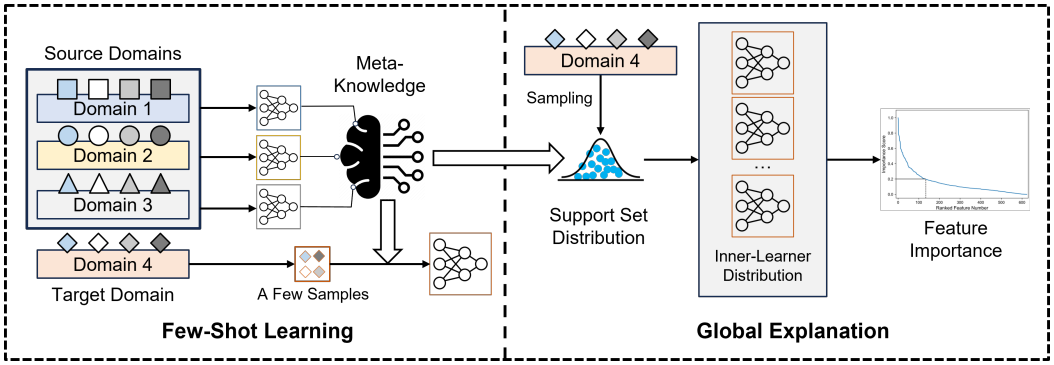

Fig. 4. Schematic overview of the XFSL framework.

4. XFSL Framework

4.1. MAML-based FSL

To enable FSL in cross-domain problems, MAML is employed in this study, where the initialization of parameters is deemed as the meta-knowledge representing the underlying characteristics of the distributions among different tasks. Let θ be the parameters for inner-learner f_θ, then the goal of MAML is to learn:

$$\omega^* = \underset{\omega}{\operatorname{argmin}} \sum_i L^{\text{meta}}(D^{(i)}_{\text{query}}; \theta^{*(i)}(\omega), \omega), \tag{1}$$

$$\text{s.t.} \quad \theta^{*(i)}(\omega) = \underset{\theta}{\operatorname{argmin}}\, L^{\text{task}}(D^{(i)}_{\text{sup}}; \theta, \omega), \tag{2}$$

where ω is initialization shared by all inner-learners; and L_{meta} and L_{task} refer to meta and task losses, respectively.

Fig. 5 illustrates the implementation of MAML for XFSL. It shows an example scenario where there are three source domains, i.e., Tasks 1–3, and one target domain Task 4. The extension to more source domains is straightforward. In meta-training, the supervised learning problem of each domain constitutes an individual task. The training process starts with initializing the inner-learner with parameters ω. For each task, a support set is sampled from the corresponding domain and used to optimize the parameters of the inner-learner. The inner-learner with updated parameters is then utilized to make predictions on the query set for the same task. The prediction error on the query set is used as a loss function to update the meta-knowledge ω. By repeating this process for all tasks, the generalization ability of the meta-knowledge ω is reinforced gradually. The final inner-learner initialized with optimal ω can quickly adapt to any similar task with a small number of training examples.

4.2. Structure of inner learner

The inner-learner in MAML is a task-wise machine learning model that is trained to fulfill a specific task. It is used to update the model parameters in a fast and efficient manner, allowing the model to adapt to new tasks quickly. The choice of the inner-learner in MAML is of vital importance as it affects the

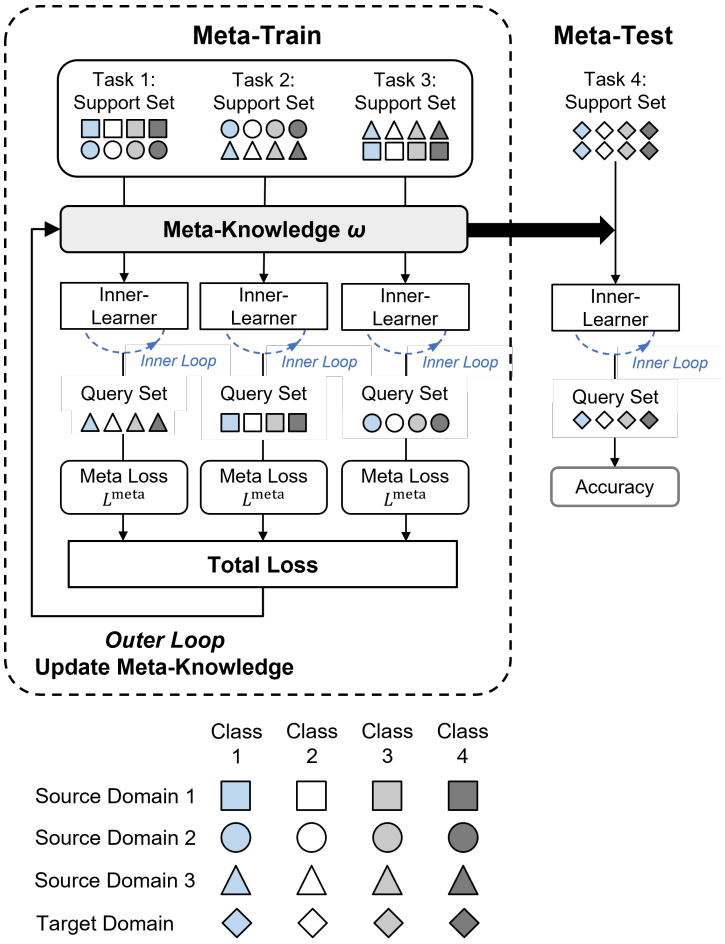

Fig. 5. The MAML procedure for XFSL.

overall performance of the meta-learning algorithm. The inner-learner should be chosen according to the following considerations:

- Task complexity: The complexity of the tasks in the meta-training phase should match the capacity of the inner-learner. For simple tasks, a simple model may be sufficient, while more complex tasks may require a more powerful model.
- Model architecture: The inner-learner should be appropriate for the type of task and data being used.

- Training data size: The inner-learner should be able to effectively learn from the limited training data that is available for each task.

- Computational resources: The inner-learner should be computationally efficient, as it will be trained many times during the meta-training phase.

According to [16], a multi-layer perceptron (MLP) model is effective in achieving high accuracy with the extracted features. Hence, to reduce the compuatational cost and avoid overfitting, MLP is adopted as the inner-learner in this work. The structure of MLP is BN $\rightarrow$ FC(624,200) $\rightarrow$ ReLU $\rightarrow$ BN $\rightarrow$ FC(200,50) $\rightarrow$ ReLU $\rightarrow$ BN $\rightarrow$ FC(50,4), where the inner learning rate is 0.01.

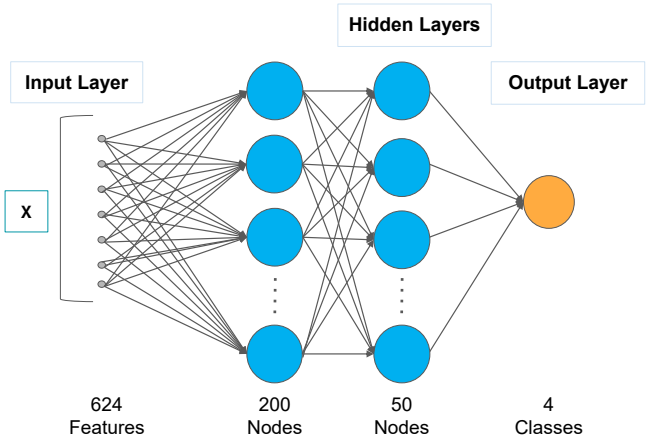

Fig. 6. Architecture of the MLP model for the inner-learner.

4.3. Optimization of inner-learner

There are mainly two loops in XFSL for FSL: inner-loop to train the inner-learner and the outer-loop to update the meta-knowledge. In the inner-loop, the inner-learner parameters θ is updated by gradient descent, minimizing the task-specific loss function on the task-specific data distribution:

$$\theta^{\text{new}} = \theta - \alpha \nabla_{\theta} L^{\text{task}}(D_{\text{sup}}^{(i)}; \theta, \omega), \tag{3}$$

where $D_{\text{sup}}^{(i)}$ is the support set for task i, $\nabla_\theta L^{\text{task}}$ is the gradient of the loss with respect to the parameters, and α is the inner-loop learning rate. The task loss is defined as

$$L^{\text{task}} = \sum_{(\boldsymbol{x_j}, \boldsymbol{y_j}) \in D_{\text{sup}}^{(i)}} l(\boldsymbol{y_j}, \hat{\boldsymbol{y}}_{\boldsymbol{j}}), \tag{4}$$

where $\boldsymbol{y_j}$ is the one-hot vector of the true label, $\hat{\boldsymbol{y}}_{\boldsymbol{j}}$ is the predicted label distribution of input $\boldsymbol{x_j}$, and l is the cross-entropy loss, i.e., for any k-dimensional probability vector t, p:

$$l(\boldsymbol{t}, \boldsymbol{p}) = -\sum_k t_k \log p_k. \tag{5}$$

One important aspect of MAML is the total number of inner-loop updates. In general, more inner-loop updates result in better performance, but also require more computational resources and time. In this study, inner-loop is chosen to be updated once for efficiency.

The query set, drawn from the same distribution with support set, is then used to evaluate the performance of the trained inner-learner, providing insights into the generalization ability of the model. Based on the average performance across all tasks, the initialization ω can be updated by

$$\omega^{\text{new}} = \omega - \sum_i \beta \nabla_\omega L^{\text{meta}}(D_{\text{query}}^{(i)}; \theta^{*(i)}(\omega), \omega), \tag{6}$$

where $\theta^{*(i)}$ is the updated parameters of inner-learner for task i and β is the meta-learning rate, which is 0.001. The meta loss is defined as

$$L^{\text{meta}} = \sum_{(\boldsymbol{x_j}, \boldsymbol{y_j}) \in D_{\text{query}}^{(i)}} l(\boldsymbol{y_j}, \hat{\boldsymbol{y}}_{\boldsymbol{j}}). \tag{7}$$

4.4. XAI for XFSL

For conventional FSL methods, the mechanism behind how the models work is difficult to interpret. To tackle this challenge, we develop an XAI method based on the saliency map method. Traditionally, the saliency map is computed by:

$$w_I^c = \frac{\partial S_c}{\partial I}, \tag{8}$$

where w_I^c is the importance value of pixel I to class c, and S_c is the output likelihood of belonging class c. The saliency map is designed for visualizing the important pixels of the image, where each pixel shares same scale and distribution. However, in our problems, the input vector is a collection of features acquired from different wavelet levels. Different features have significant different scales and distributions. To address this problem, we compute the saliency map w.r.t the output of first batchnorm layer. Therefore,

$$w_i^c = \frac{\partial S_c}{\partial x_i} \cdot \sigma_{x_i}, \tag{9}$$

where w_i^c is the importance value of input feature x_i to class c, and S_c is the probability that an instance is assigned to class c.

Conventional saliency maps have been widely recognized as local methods, offering insights into feature importance at the individual instance level [41]. Therefore, it is difficult to draw general and global conclusions about feature importance across an entire dataset or model. Moreover, the saliency map is sensitive to small perturbations or changes in the input data or the model parameters, leading to potentially inconsistent interpretations.

To alleviate these limitations, we propose using the meta-knowledge learned from XFSL to establish the saliency map. Suppose Ω is the meta-knowledge, then the importance of a feature for class c is computed by

$$w_i^c = \mathbb{E}_{D_{\text{sup}} \in \mathscr{D}_{\text{sup}}} \mathbb{E}_{\boldsymbol{x}^q \in D - D_{\text{sup}}} \frac{\partial S_c(\boldsymbol{x}^q; D_{\text{sup}}, \Omega)}{\partial x_i^q} \cdot \sigma_{D_{\text{sup}}, i}, \tag{10}$$

where w_i^c is the importance value of ith input feature to class c, D is the entire dataset, $\mathscr{D}_{\text{sup}}$ is the collection of the 1-shot support set D_{sup}, x_i^q represents the ith feature of $\boldsymbol{x}^q$, $\sigma_{D_{\text{sup}},i}$ is standard deviation of ith feature in D_{sup}.

Unlike the conventional saliency map, this method incorporates the meta-knowledge about the learning process across multiple tasks, which establishes a global context that encompasses a diverse range of data instances and variations. Given meta-knowledge Ω, the w_i^c is fixed, providing a global explanation of the class c. As such, this XAI method can effectively leverage the meta-knowledge to establish a global model explanations, thereby enhancing the ability of saliency

map. Furthermore, we can use this method to design an effective feature reduction procedure, which will be reported in Section 5.3.

The psedo-code for XFSL is formally presented in Algorithm 1.

5. Results and Discussion

This section examines the effectiveness of the proposed XFSL method using real-world UMW datasets. Section 5.1 presents the experimental design and data collection process. Sections 5.2 evaluates the performance of XFSL and baseline methods with several case studies. The XAI results are reported in Section 5.3.

5.1. Dataset description

Table 1: Design of experiments and domain group M configuration: combinations of four materials along with different welding time and four tool conditions.

Domain	Material	Welding Time	Dataset Size	Classification Goal			
CC	Cu-Cu	0.9 s	200	TC0	TC1	TC2	TC3
AC	Al-Cu	0.5 s	200				
CA	Cu-Al	0.5 s	200				
AA	Al-Al	0.9 s	200				

Domain group M: The top and bottom workpieces are made of 50.8 mm (length) $\times$ 25.4 mm (width) $\times$ 0.25 mm (thickness) Cu or Al sheet with different welding time. Welding samples are generated on different materials with welding time combinations, as shown in Table 1. The sampling frequency for signal acquisition is 250 kHz. It is worth noting that Cu-Al (CA) and Al-Cu (AC) are different domains since the vibrational patterns during welding are different. Each domain consists of 200 samples that are evenly distributed over four tool conditions: new horn/new anvil (TC0), new horn/worn anvil (TC1), worn horn/new anvil (TC2), and worn horn/worn anvil (TC3). Therefore, 800 samples are collected in total.

Algorithm 1 XFSL

Require: Source domains D^1, D^2, ..., D^k, target domain D_{target}

Require: Learning rate α, β, number of shot k_s, the class of interest c for importance estimation of jth feature

// **Step 1: Domain Split**

Split the source domain D^i into training domain D^i_{tr} and validation domain D^i_{va}

// **Step 2: Training Meta-Learning Model**

Randomly initialize ω

while ω has not converged **do**

  for i = 1,2,...,k **do**

    Sample k_s-shot support set $D^{(i)}_{\text{sup}}$ and query set $D^{(i)}_{\text{query}}$ from D^i_{tr}

    $\theta = \omega$ // Initialize θ of inner learner f_θ with ω

    // Train the inner-learner with support set. We update θ once for efficiency

    $\theta^{*(i)} \leftarrow \theta - \alpha \nabla_\theta L^{\text{task}}(D^{(i)}_{\text{sup}}; \theta, \omega)$

  end for

  // Update the common initialization ω

  $\omega \leftarrow \omega - \sum_i \beta \nabla_\omega L^{\text{meta}}(D^{(i)}_{\text{query}}; \theta^{*(i)}, \omega)$

  Repeat 6-12 on D^i_{va} to get the validation accuracy

end while

Ω is the ω with the highest training and validation accuracy

Train an inner-learner with support set on D_{target} and evaluate the performance

// **Step 3: Obtaining the Global Feature Importance of feature** j **for Class** c

$w^c_j = 0$ // Initialize w^c_j

for i = 1,2,...,$|\mathscr{D}_{\text{sup}}|$ **do**

  Sample support set D_{sup} and calculate $\sigma_{D_{\text{sup}},i}$

Algorithm 1 XFSL (continued)

Train and obtain the inner-learner with Ω and D_{sup}

Get the class probability function S_c from the inner-learner

for Go over all $\boldsymbol{x}^q$ sampled from $|D - D_{\text{sup}}|$ **do**

$w_j^c \leftarrow w_j^c + \frac{\partial S_c(\boldsymbol{x}^q; D_{\text{sup}}, \Omega)}{\partial x_j^q} \cdot \sigma_{D_{\text{sup}},j}$

end for

end for

Table 2: Design of experiments and domain group S configuration.

Domain	Dataset Size	Classification Goal		
Clean	90			
Polished	90	New	Worn	DMGD
Contam	90			

Domain group S: Both the top and bottom workpieces are made of 50.8 mm (length) × 25.4 mm (width) × 0.20 mm (thickness) Cu sheets. The welding time is selected as 1.0 s and unlike the Domain group M, the sampling frequency of Domain Group S is 200 kHz. As shown in Table 2, the surface conditions in different domains are different, including "clean" surface condition, "polished" surface condition and "contaminated" surface condition. A "clean" surface condition is cleaned using alcoholic wipes prior to welding. The contact faces of the workpieces polished with sandpaper before welding is defined as a "polished" surface condition. A "contaminated" surface condition is obtained by applying one drop of cutting fluid to the workpiece surface. Each surface condition consists of 90 samples that are evenly distributed for three tool conditions, resulting in 270 samples in total. The goal of the S domains is to classify three tool conditions: new horn/new anvil (New), worn horn/worn anvil (Worn) and damaged horn/damaged anvil (DMGD). It is worth noting that despite that the tool labeled "New" and "Worn" are also used for domain group M, the "DMGD" condition is not used in the M domains.

Domain group T: The data in this domain group are identical to domain

group S. The difference is that the goal of domain group T is to classify the surface conditions: clean, polished and contaminated. Different domains represent different tool conditions, as shown in Table 3. Those domains mimic a scenario where predicting the workpiece surface condition is desired while the welding tools are fixed.

Table 3: Design of experiments and domain group T configuration.

Domain	dataset size	Classification Goal		
DMGD	90			
New	90	Clean	Polished	Contaminated
Worn	90			

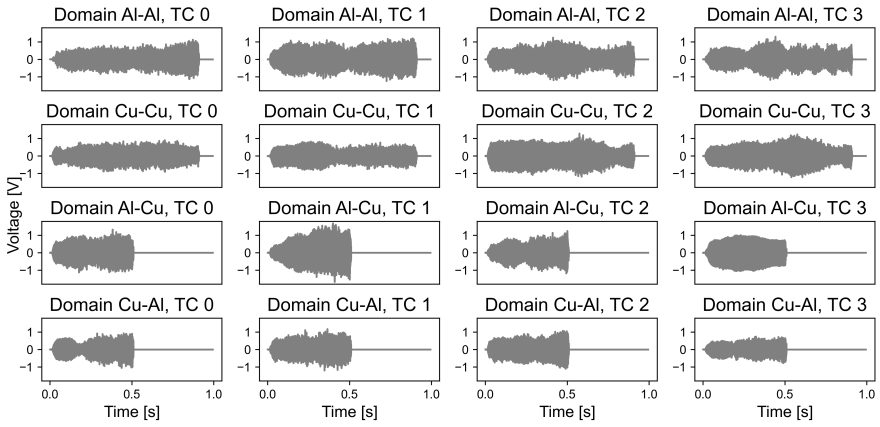

Fig. 7. Comparison of microphone signals in different M domains.

Fig. 7 compares some typical microphone signals in each M domain. The data is highly different in terms of patterns and length in different domains. Fig. 8 shows the comparison of feature entropy of level 0 between Domain Cu-Cu and Domain Al-Cu. It is seen that even after feature extraction, the distribution of different tool conditions in the feature space is still significantly different among domains, which highlights the challenges in effective knowledge transfer.

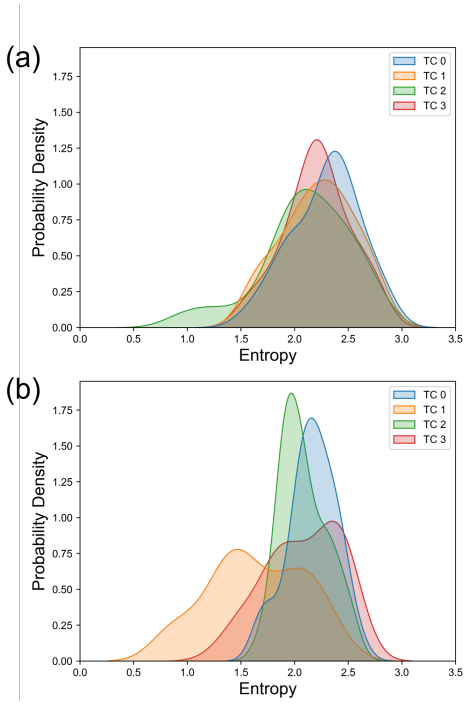

Fig. 8. Comparison of distribution of entropy at level 0 (one of the extracted features) between (a) Domain Cu-Cu and (b) Domain Al-Cu. The distribution is estimated by kernel density estimation.

5.2. FSL results

The performance of the proposed method is evaluated on three FSL problems, each of which focuses on one domain group (M, S or T). A neural network (NN) with identical structure to the inner-learner of XFSL is built as the baseline model. All models are run on a desktop with Intel i7-12700H CPU @ 2.30 GHz and Nvidia GeForce RTX 3060 GPU. The training time of XFSL on all domain group problems is within 5 minutes for one run.

For domain group M, the model performance is evaluated on four domains. We pick three domains in turn as the source domains and use the remaining one as the target domain. Each source domain dataset is further divided into meta-validation set and meta-train set with the ratio of 1:4. Therefore, the target domain, training source domain and validation source domain are 200,

480, and 120 samples, respectively.

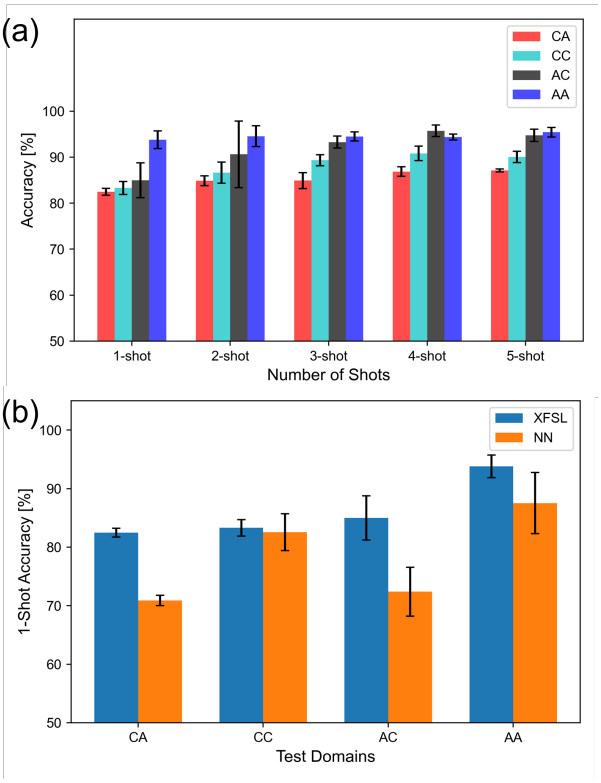

Fig. 9. The model performance for domain group M. (a) Performance of XFSL with different numbers of shots. (b) Performance comparison of XFSL and NN.

Fig. 9 displays the model performance for domain group M. As shown in Fig. 9(a), the XFSL model achieves good performance in all FSL problems. Specifically, the model in 1-shot problem on target domain AA can reach 90% accuracy. Even the 1-shot accuracy on target domain CA, which is lowest among all problems, is over 80%. Moreover, as the number of shots increases, the accuracy consistently increases and then converges when it reaches 4- or 5-shot. Meanwhile, the variation in classification accuracy decreases as the number of shot increases, indicating improvement in robustness. This phenomenon arises mainly because a small number of shots leads to imprecise parameter estimation of the batchnorm layer of the inner-learner, especially for 1- and 2-shot cases.

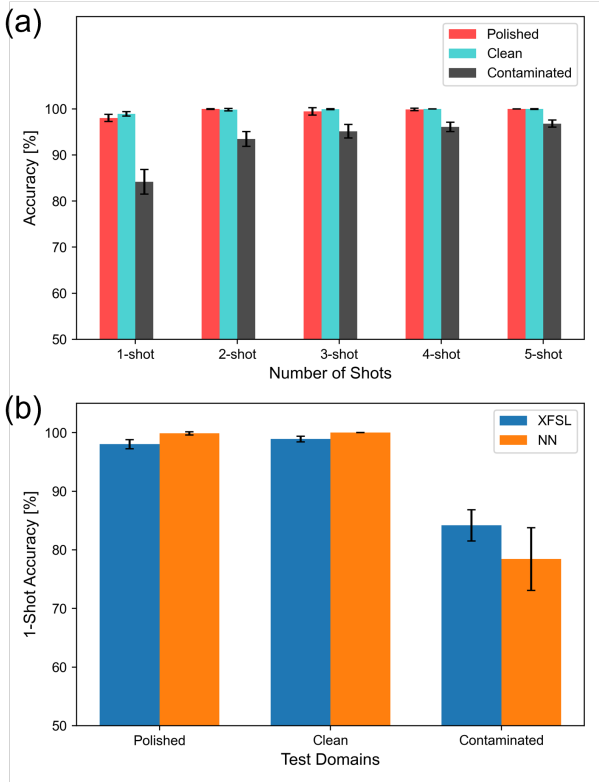

Fig. 10. The model performance for domain group S. (a) Performance of XFSL with different numbers of shots. (b) Performance comparison of XFSL and NN.

This problem is greatly alleviated as the number of shots increases, thereby improving the consistency and robustness of the model.

The comparative results for XFSL and NN are shown in Fig. 9(b). It is seen that XFSL, with an average accuracy of 82.48% on target domains, has superior performance over the baseline method NN, especially on {AC, CC, AA→CA} with an improvement of 11.58%. Moreover, NN has much larger variations in classification performance. Therefore, the proposed model outperforms NN in both accuracy and robustness.

The proposed model performance for domain group S is shown in Fig. 10. It is seen from Fig. 10(a) that TCM for "Polished" and "Clean" workpiece surface conditions, XFSL achieves close-to-perfect performance even with 1-

shot. TCM for the "Contaminated" domain is more challenging. However, the prediction accuracy of XFSL improves when more data becomes available. As shown in Fig. 10(b), NN has comparable performance with XFSL and both models achieved an accuracy of close to 100% for test domains of "Polished" and "Clean." The performance of NN on the "Contaminated" domain is much worse than XFSL. This could be attributed to the fact that the data distributions of "Clean" and "Polish" domains are more similar with each other, while the "Contaminated" domain differs from these two domains significantly. XFSL is able to capture such heterogeneity and achieves higher classification accuracy than NN.

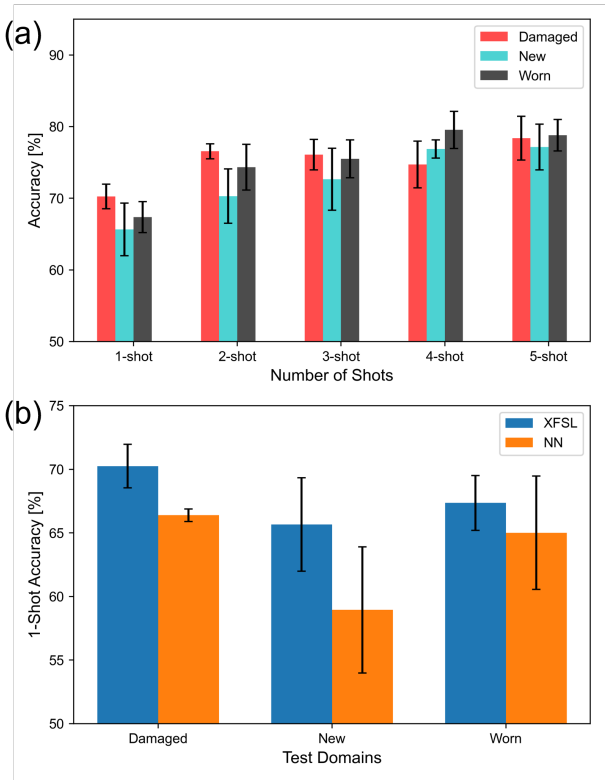

Fig. 11. The model performance for domain group T. (a) Performance of XFSL with different numbers of shots. (b) Performance comparison of XFSL and NN.

Fig. 11 shows the results for domain group T. Different results than domain

groups M and S are observed. The accuracy for classifying workpiece surface conditions ranges from 65% to 75%. As shown by Fig. 11(a), XFSL generally benefits from having access to more labeled data in the target domain. The "Worn" domain has lower accuracy than the "New" and "Damaged" domains, indicating that classifying workpiece surface conditions with worn tools is more challenging than other tool conditions. Fig. 11(b) shows that XFSL outperforms neural networks. It is also seen that the XFSL models can consistently deliver over 65% accuracy. Comparing domain group T with M and S, we can see that classifying workpiece conditions is much more challenging than TCM.

5.3. XAI for FSL

So far, our XFSL models have shown to be successful in extracting the meta-knowledge for different types of process anomaly detection. However, it remains unclear that which features, frequencies, sensors or feature categories predominantly contribute to the model's performance. This ambiguity poses a substantial challenge to the interpretation of the results and feature reduction to reduce pre-processing and inference cost. Therefore, this section will demonstrate the ability of XFSL in identifying important features. As one example, we focus on Domain Group M and {CA,CC,AC→AA} scenario in this study.

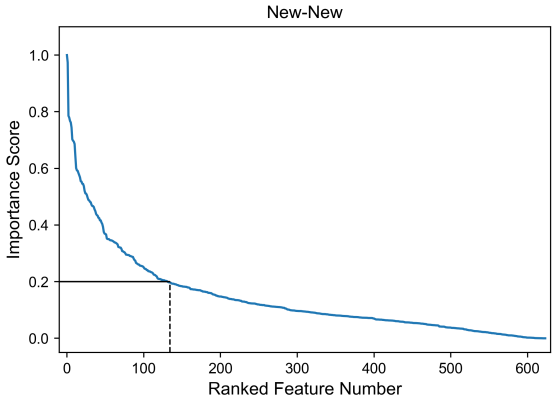

Fig. 12. Feature importance scores ranked from highest to lowest for identifying New-New condition in {CA,CC,AC→ AA} task.

Fig. 12 shows a representative example of feature importance scores. It is noteworthy that only a very small portion of the features have high importance scores (exceeding 0.2). This implies that the majority of features can be deemed as redundant for the classification task, thereby suggesting the possibility of their removal to streamline the model without significantly impacting its performance.

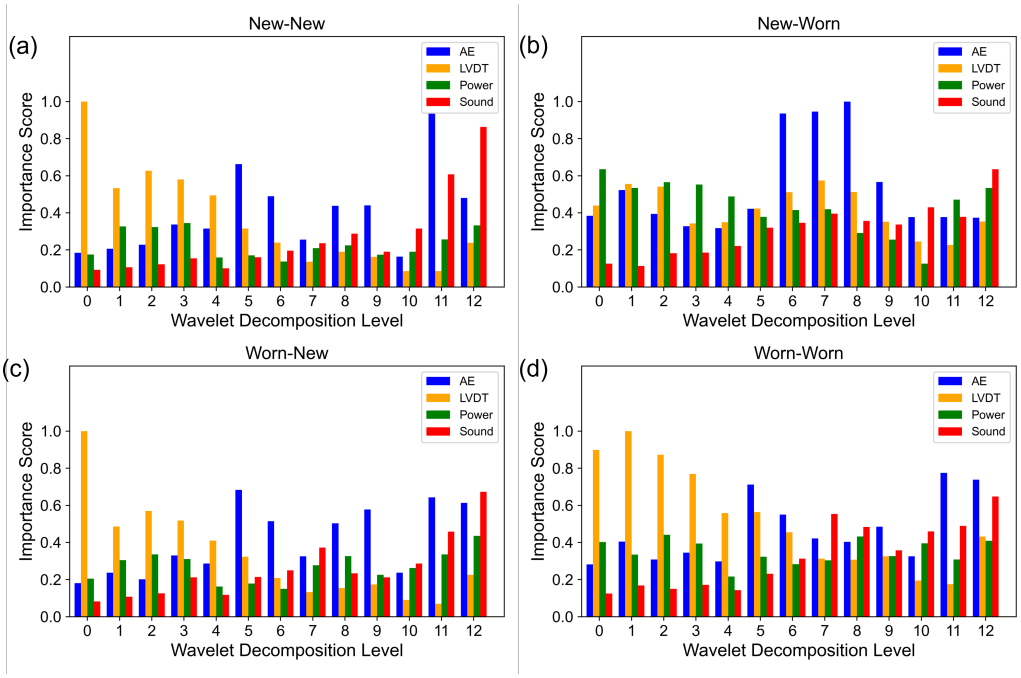

Fig. 13. The importance scores of different level of wavelet coefficients, where each level corresponds to a frequency band. A higher level represents a lower frequency band.

In this study, the original signal is decomposed into different levels of wavelet coefficients, where each level corresponds to a frequency band. Higher levels correspond to lower frequencies. Therefore, by aggregating the scores of all features corresponding to the same level of wavelet and normalizing the results, the comparison of importance for different frequencies can be obtained, as shown in Fig. 13. It is revealed that high-frequency bands of LVDT signals are more useful than low-frequency bands. Conversely, lower frequency bands of sound

signals play a more important role. This difference may be attributed to the inherent characteristics of these sensors: LVDT sensors effectively capture high-frequency information, whereas sound sensors are more attuned to low-frequency details. Table 4 compares the overall importance of each sensor. It is seen that AE and LVDT sensors contribute more significantly to the overall prediction, whereas power and sound sensors play less decisive roles.

Table 4: Sensor importance scores for 1-shot learning in Domain AA.

	AE	LVDT	Power	Sound
New-New	5.13	4.69	3.02	3.44
New-Worn	6.94	5.42	5.66	4.03
Worn-New	5.33	4.36	3.51	3.35
Worn-Worn	6.05	6.86	4.57	4.29
Average	5.86	5.33	4.19	3.78

Fig. 14 shows the importance scores of different categories of features towards each label. It is revealed that the entropy and the zero-crossing rate, capturing fundamental characteristics of the signal, emerge as the most crucial features. In contrast, the median and mean are found to contribute minimally to the prediction, suggesting they could be excluded from the feature set without substantially degrading classification performance.

The prior analysis shows that only a small portion of the features exhibit high importance scores. This finding suggests a good opportunity for feature selection based on the importance scores. Here, we devise the following feature selection strategy:

1. Suppose we want to extract r features, and there are n classes in total. Initialize the feature set S_r with the empty set.

2. For the first class, collect the feature (or features of same frequency/category) with the highest importance score to form set S^1.

3. Repeat Step 2 on the remaining classes until all classes are done. Then

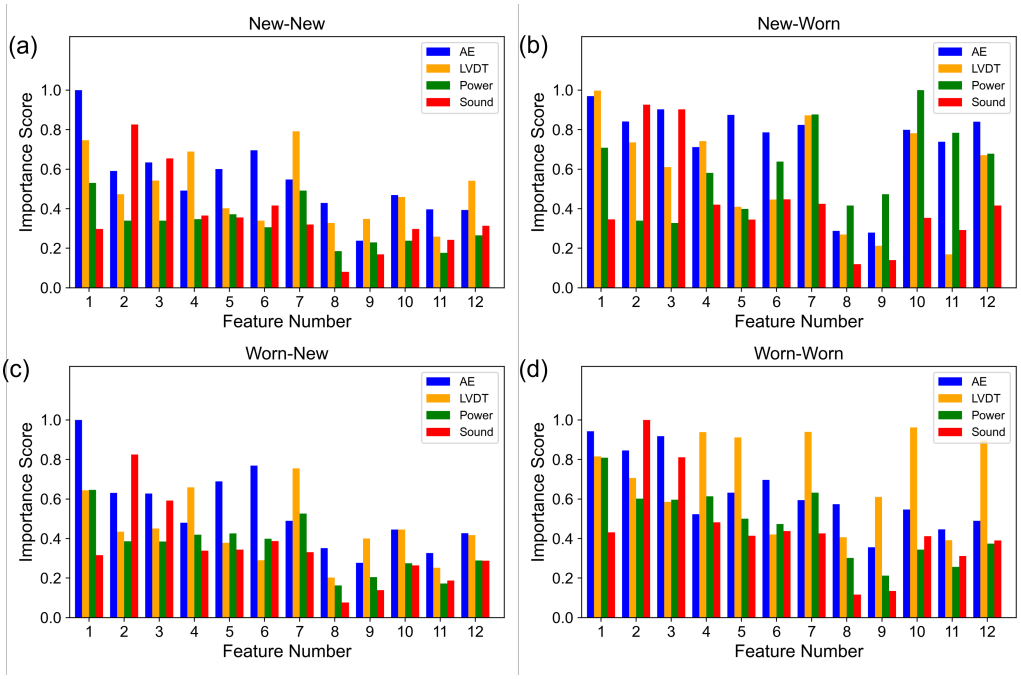

Fig. 14. The importance scores of different feature categories, where 1: entropy; 2: zero crossing rate; 3: mean crossing rate; 4-7: 5th/25th/75th/95th percentile; 8: median; 9: mean; 10: standard deviation; 11: variance; 12: root mean square value.

update S_r with $S_r \bigcup S^1 \bigcup S^2 \cdots \bigcup S^n$.

4. Repeat Steps 2 and 3 with the second highest ones until the size of S_r reaches r.

We use the strategy to select a quarter (156 features) and a half (312 features) of the original features based on their importance scores obtained from {CA,CC,AC→AA}. Subsequently, we perform an evaluation of the refined feature set on 1-shot learning tasks. To ensure the robustness and validity of the feature selection by XAI for meta-learning, the performance assessment is conducted across all four tasks, each employing the identically selected feature set. This multi-task evaluation serves to substantiate the generalizability and efficacy of our feature selection strategy.

The results are reported in Table 5. It is shown that retaining half or even

a quarter of input features, the model still sustains high accuracy for 1-shot problem, which underscores not only the efficacy but also the generalizability of the XAI method. Interestingly, in some scenarios, such as {AC,AA,CA→CC}, reducing the number of input features can even boost the performance. This is because for 1-shot learning problems, some redundant features may bring significant disturbances to the classification model due to extremely small number of training samples. Removal of the non-informative features can increase the model performance. Another important finding is that all feature selection methods can achieve comparable performance with the original one, providing a flexibility in choosing the methods depending on the needs.

Table 5: Comparison of accuracy after feature reduction. Note that the feature is selected based on the meta-knowledge learned from {CA,CC,AC→AA}.

Number of Features	Features Selection	Test Domain			
		CA	CC	AC	AA
624	None	81.49%	83.89%	87.50%	96.79%
312	By Feature	85.57%	88.51%	95.56%	97.11%
156	By Feature	89.96%	90.15%	93.82%	96.60%
312	By Frequency	82.23%	90.44%	89.23%	98.77%
156	By Frequency	79.06%	87.89%	80.80%	98.38%
312	By Category	87.76%	76.09%	80.41%	93.03%
156	By Category	85.74%	75.04%	87.35%	88.09%

6. Conclusion

This research proposes a novel XFSL framework to efficiently detect process anomalies online in UMW. The framework is designed to effectively learn and transfer the meta-knowledge across multiple process configurations (domains), thereby enhancing the generalization capability of the learning model in data-scarce situations. A series of case studies that simulate different real-world applications are conducted to validate the performance of the XFSL framework.

These case studies span a broad range of TCM scenarios, including various material types, surface conditions, and welding settings, mimicking the heterogeneity and complexity typically encountered in industrial environments. The results show that XFSL outperforms baseline methods in all case studies, demonstrating the power of XFSL in data-scarce situations.

By leveraging the learned meta-knowledge, XFSL can further enable the explainability of the model and evaluation of feature importance. This method not only enhances the reliability of the saliency map but also expands its interpretability scope from individual instances to the overarching patterns in the meta-data, thereby bridging the gap between local and global model interpretability. A feature selection strategy is devised on the basis of our XAI method. Quantitative results on multiple learning tasks show that the models trained on the reduced feature set consistently perform as well as those trained on the full feature set, and sometimes achieve better performance. Therefore, the proposed XAI method enables a robust and comprehensive interpretation of feature importance across multiple learning tasks.

The results of this work suggest two future research directions. First, the ability of knowledge transfer across domain groups (M, S, and T) is highly desirable to further extend the FSL capabilities. Development of this ability will need to address the heterogeneity in class types and the numbers of classes. Second, domain generalization [42], which assumes that data is absolutely unavailable in the target domain, will be able to unlock more powerful predictive capabilities (e.g., before production launch in manufacturing). It will require a more effective method for extracting fundamental meta-knowledge that is invariant across domains.

7. Acknowledgments

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

References

[1] W. W. Cai, B. Kang, S. J. Hu, Ultrasonic Welding of Lithium-Ion Batteries, ASME Press, 2017. doi:https://doi.org/10.1115/1.861257.

[2] X. M. Cheng, K. Yang, J. Wang, W. T. Xiao, S. S. Huang, Ultrasonic system and ultrasonic metal welding performance: A status review, Journal of Manufacturing Processes 84 (2022) 1196–1216. doi:https://doi.org/10.1016/j.jmapro.2022.10.067.

[3] A. Siddiq, E. Ghassemieh, Thermomechanical analyses of ultrasonic welding process using thermal and acoustic softening effects, Mechanics of Materials 40 (12) (2008) 982–1000. doi:https://doi.org/10.1016/j.mechmat.2008.06.004.

[4] 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.

[5] Z. Ni, F. Ye, Ultrasonic spot welding of aluminum alloys: A review, Journal of Manufacturing Processes 35 (2018) 580–594. doi:https://doi.org/10.1016/j.jmapro.2018.09.009.

[6] T. H. Kim, J. Yum, S. J. Hu, J. P. Spicer, J. A. Abell, Process robustness of single lap ultrasonic welding of thin, dissimilar materials, CIRP annals 60 (1) (2011) 17–20. doi:https://doi.org/10.1016/j.cirp.2011.03.016.

[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. doi:https://doi.org/10.1016/j.jmsy.2018.04.014.

[8] 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. doi:10.1109/TADVP.2008.2009927.

[9] 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. doi:10.3934/mbe.2020379.

[10] 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. doi:https://doi.org/10.1016/j.ymssp.2022.109473.

[11] 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. doi:https://doi.org/10.1016/j.jmsy.2013.05.006.

[12] 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) (2015) 051005. doi:10.1115/1.4031677.

[13] 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. doi:https://doi.org/10.1016/j.jmsy.2020.06.020.

[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. doi:https://doi.org/10.1016/j.jmapro.2020.12.050.

[15] 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) (2014) 051019. doi:10.1115/1.4028059.

[16] K.-C. Lu, Y. Meng, Z. Dong, C. Shao, Online cost-effective classification of mixed tool and material conditions in ultrasonic metal welding: Towards integrated monitoring and control, in: International Manufacturing Science and Engineering Conference, Vol. Volume 2: Manufacturing Equipment and Automation; Manufacturing Processes; Manufacturing Systems; Nano/Micro/Meso Manufacturing; Quality and Reliability, 2023, p. V002T07A007. doi:10.1115/MSEC2023-104868.

[17] 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. doi:https://doi.org/10.1016/j.jmsy.2016.01.001.

[18] X. Shi, S. Yu, L. Li, J. Zhao, Anvil state identification based on acceleration signals in ultrasonic metal welding of lithium batteries, Journal of Manufacturing Processes 70 (2021) 67–77. doi:https://doi.org/10.1016/j.jmapro.2021.08.023.

[19] 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. doi:https://doi.org/10.1016/j.jmapro.2022.02.057.

[20] Y. Wu, Y. Meng, C. Shao, End-to-end online quality prediction for ultrasonic metal welding using sensor fusion and deep learning, Journal of Manufacturing Processes 83 (2022) 685–694. doi:https://doi.org/10.1016/j.jmapro.2022.09.011.

[21] S. J. Hu, Evolving paradigms of manufacturing: From mass production to mass customization and personalization, Procedia CIRP 7 (2013) 3–8,

forty Sixth CIRP Conference on Manufacturing Systems 2013. doi:https://doi.org/10.1016/j.procir.2013.05.002.

[22] C. Tan, S. J. Hu, H. Chung, K. Barton, C. Piya, K. Ramani, M. Banu, Product personalization enabled by assembly architecture and cyber physical systems, CIRP Annals 66 (1) (2017) 33–36. doi:https://doi.org/10.1016/j.cirp.2017.04.106.

[23] P. Wang, R. X. Gao, Transfer learning for enhanced machine fault diagnosis in manufacturing, CIRP Annals 69 (1) (2020) 413–416. doi:https://doi.org/10.1016/j.cirp.2020.04.074.

[24] W. Jiao, Q. Wang, Y. Cheng, Y. Zhang, End-to-end prediction of weld penetration: A deep learning and transfer learning based method, Journal of Manufacturing Processes 63 (2020) 191–197. doi:10.1016/j.jmapro.2020.01.044.

[25] 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. doi:https://doi.org/10.1016/j.jmsy.2020.12.009.

[26] M. Mehta, S. Chen, H. Tang, C. Shao, A federated learning approach to mixed fault diagnosis in rotating machinery, Journal of Manufacturing Systems 68 (2023) 687–694. doi:https://doi.org/10.1016/j.jmsy.2023.05.012.

[27] Y. Wang, Q. Yao, J. T. Kwok, L. M. Ni, Generalizing from a few examples: A survey on few-shot learning, ACM Comput. Surv. 53 (3) (2020) 1–34. doi:10.1145/3386252.

[28] C. Finn, P. Abbeel, S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, in: International conference on machine learning, PMLR, 2017, pp. 1126–1135.

[29] Z. Li, M. Kumar, W. Headden, B. Yin, Y. Wei, Y. Zhang, Q. Yang, Learn to cross-lingual transfer with meta graph learning across heterogeneous languages, in: Conference on Empirical Methods in Natural Language Processing, 2020.

[30] Q. Dou, D. Coelho de Castro, K. Kamnitsas, B. Glocker, Domain generalization via model-agnostic learning of semantic features, in: Advances in Neural Information Processing Systems, 2019.

[31] C. Liu, Y. Li, J. Li, J. Hua, A meta-invariant feature space method for accurate tool wear prediction under cross conditions, IEEE Transactions on Industrial Informatics 18 (2) (2022) 922–931. doi:10.1109/TII.2021.3070109.

[32] D. Wang, Q. Liu, D. Wu, L. Wang, Meta domain generalization for smart manufacturing: Tool wear prediction with small data, Journal of Manufacturing Systems 62 (2022) 441–449. doi:https://doi.org/10.1016/j.jmsy.2021.12.009.

[33] Y. Zhou, S. Booth, M. T. Ribeiro, J. Shah, Do feature attribution methods correctly attribute features?, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 36, 2022, pp. 9623–9633. doi:10.1609/aaai.v36i9.21196.

[34] K. Simonyan, A. Vedaldi, A. Zisserman, Deep inside convolutional networks: Visualising image classification models and saliency maps, arXiv preprint arXiv:1312.6034.

[35] M. D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: Computer Vision – ECCV 2014, Springer International Publishing, Cham, 2014, pp. 818–833.

[36] Y. Su, P. Yan, R. Yi, J. Chen, J. Hu, C. Wen, A cascaded combination method for defect detection of metal gear end-face, Journal of Manufactur-

ing Systems 63 (2022) 439–453. doi:https://doi.org/10.1016/j.jmsy.2022.05.001.

[37] Z. Ye, J. Yu, Aksnet: A novel convolutional neural network with adaptive kernel width and sparse regularization for machinery fault diagnosis, Journal of Manufacturing Systems 59 (2021) 467–480. doi:https://doi.org/10.1016/j.jmsy.2021.03.022.

[38] 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.

[39] Y. Cheng, R. Yu, Q. Zhou, H. Chen, W. Yuan, Y. Zhang, Real-time sensing of gas metal arc welding process–a literature review and analysis, Journal of Manufacturing Processes 70 (2021) 452–469.

[40] R. Yu, Y. Cao, H. Chen, Q. Ye, Y. Zhang, Deep learning based real-time and in-situ monitoring of weld penetration: Where we are and what are needed revolutionary solutions?, Journal of Manufacturing Processes 93 (2023) 15–46. doi:https://doi.org/10.1016/j.jmapro.2023.03.011.

[41] A. Das, P. Rad, Opportunities and challenges in explainable artificial intelligence (xai): A survey, arXiv preprint arXiv:2006.11371.

[42] K. Zhou, Z. Liu, Y. Qiao, T. Xiang, C. C. Loy, Domain generalization: A survey, IEEE Transactions on Pattern Analysis and Machine Intelligence (2022) 1–20doi:10.1109/tpami.2022.3195549.