

Nondestructive Fatigue Life Prediction for Additively Manufactured Metal Parts through a Multimodal Transfer Learning Framework

Abstract

Understanding the fatigue behavior and accurately predicting the fatigue life of laser powder bed fusion (L-PBF) parts remain a pressing challenge due to complex failure mechanisms, time-consuming tests, and limited fatigue data. This study proposes a physics-informed data-driven framework, a multimodal transfer learning (MMTL) framework, to understand process-defect-fatigue relationships in L-PBF by integrating various modalities of fatigue performance, including process parameters, XCT-inspected defects, and fatigue test conditions. It aims to leverage a pre-trained model with abundant process and defect data in the source task to predict fatigue life nondestructively with limited fatigue test data in the target task. MMTL employs a hierarchical graph convolutional network (HGCN) to classify defects in the source task by representing process parameters and defect features in graphs, thereby enhancing its interpretability. The feature embedding learned from HGCN is then transferred to fatigue life modeling in neural network layers, enabling fatigue life prediction for L-PBF parts with limited data. MMTL validation through a numerical simulation and real-case study demonstrates its effectiveness, achieving an F1-score of 0.9593 in defect classification and a mean absolute percentage log error of 0.0425 in fatigue life prediction. MMTL can be extended to other applications with multiple modalities and limited data.

Keywords: Laser powder bed fusion; Process-defect-fatigue relationships; Fatigue life prediction; Defect classification; Multimodal transfer learning; Hierarchical graph convolutional network.

1 Introduction

Understanding the fatigue performance of additively manufactured parts and assessing their fatigue life are critical for the further adoption of additive manufacturing (AM), especially laser powder bed fusion (L-PBF), into various engineering applications under cyclic loading. Currently, fatigue failure is estimated to account for approximately 90% of mechanical failures in metallic structures (Wei et al. (2022); Peng et al. (2022)). The empirical method to assess the fatigue life of L-PBF parts based on fatigue testing is destructive and time-consuming. The test usually takes days or weeks to break one specimen, and it is unfeasible to test a large number of specimens. Moreover, even fabricated under the same conditions, parts can exhibit significant differences in microstructure, defects, and properties due to complex local process dynamics (e.g., laser-powder interactions). This variability leads to a large scatter in fatigue performance, making it impractical to fit an accurate distribution of

fatigue life with a limited number of tested specimens and predict the fatigue life for untested specimens. Recently, X-ray computed tomography (XCT) has emerged as a nondestructive inspection method in AM (Kim et al. (2019)), having great potential to enable fatigue life prediction for each individual L-PBF part.

The objective of this work is to model and predict the fatigue life of L-PBF parts from nondestructive XCT inspection by integrating physics knowledge of crack initiation and fatigue failure in L-PBF. Three modeling assumptions are derived from research experiments and domain knowledge:

(1) Process-induced volumetric defects are the most critical mechanisms influencing the fatigue life of L-PBF parts. Serving as stress risers, volumetric defects in L-PBF parts can initiate cracks under cyclic loading and thereby compromise fatigue performance in various materials (Meneghetti et al. (2019); Xu et al. (2023)). Large volumetric defects primarily determine the mechanical properties (Peng et al. (2022)), and they dominate the effect of microstructure anisotropy when their sizes are eight times the width of grains (Hu et al. (2020)). While post-process treatments can reduce the number of defects in L-PBF parts, they cannot fully eliminate them, particularly large or irregular defects (Pegues et al. (2020); Shao et al. (2017)). Other major issues, such as surface roughness, could be mitigated through polishing or machining.

(2) Different types of volumetric defects are generated under different energy inputs, which have distinct impacts on stress concentration and crack initiation due to their size and morphology. Volumetric defects in L-PBF parts, such as keyholes (KHs), lack of fusions (LOFs), and gas-entrapped pores (GEPs) (Poudel et al. (2022)), are generated under different energy inputs controlled by process parameters such as laser power and scanning speed, as shown in Figure 1 (a). KHs, with large and round shapes, are induced in unstable keyhole-shaped melt pools due to excessive energy input, acting as stress concentration points that accelerate crack initiation and propagation, thereby reducing fatigue strength. GEPs, small and highly spherical, result from the entrapment of shield gas or gas pores in powder particles and are randomly distributed, having less impact on fatigue performance (Poudel et al. (2022)). LOFs, with irregular shapes, are generated by insufficient energy input, leading to significant scatter in fatigue life (Li et al. (2022)). They can co-occur under the same process parameters due to local process variation. Classifying

different types of defects is valuable to learning the relationship between defects and their stress concentration, which will contribute to fatigue modeling and prediction with limited fatigue testing specimens.

(3) The critical defect that initiates crack and leads to fatigue failure of the test specimen is among all the defects detected by XCT nondestructively. XCT scans can detect thousands of volumetric defects in a small scanning area of L-PBF specimens, allowing for the characterization of their size and morphology (Du Plessis et al., 2020), as illustrated in Figure 1 (b). Critical defect, which has the largest stress concentration under fatigue loading and initiates cracks, is easy to detect, especially for KHs and LOFs, as discussed by Poudel et al. (2022). It is inspected from the fractography once the test specimen is broken, as shown in Figure 1 (c). Therefore, its defect features from XCT and fractography can be related to the tested fatigue life.

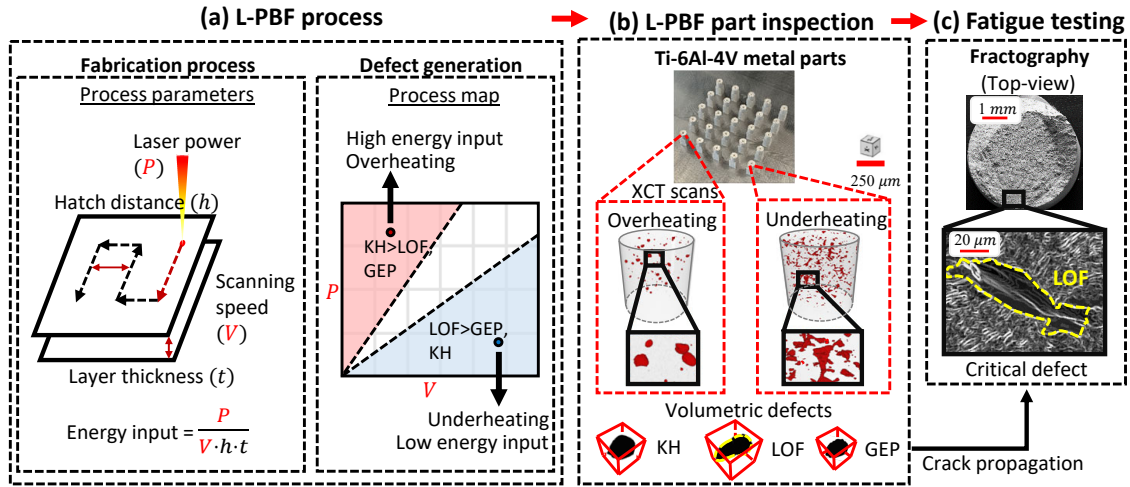


Figure 1: A process flow of L-PBF fatigue performance research from fabrication, process optimization, nondestructive inspection using XCT, and fatigue testing with fractography. (a) The energy input in fabrication is controlled by process parameters, leading to the co-occurrence of KHs, LOFs, and GEPs in one part. (b) XCT scans reveal the shapes of volumetric defects. (c) Fractography, after fatigue testing, identifies critical defects, such as LOFs, which initiate cracks and cause L-PBF part fatigue failure.

Based on these three research assumptions, we propose a multimodal transfer learning (MMTL) framework to incorporate different modalities of fatigue performance, such as L-PBF process parameters, XCT inspected defects, and fatigue test conditions, into a transfer learning (TL) framework. It aims to leverage the feature embedding (i.e., transferred knowl-

edge) learned from abundant XCT-scanned volumetric defects in defect classification and transfer it into fatigue life prediction with limited test specimens. This study includes a numerical simulation and a real-case study to quantify the relationships between the L-PBF process, volumetric defects, fatigue tests, and performance, discovering insights about fatigue performances of L-PBF Ti-6Al-4V parts. Such insights will help unravel the impact of the volumetric defect on crack initiation (Qin et al. (2021)), facilitate optimization of the L-PBF manufacturing process, and promote the further adoption of L-PBF in engineering applications. Additionally, MMTL can be applied to fatigue life prediction of L-PBF parts across different materials since defects caused by L-PBF process dynamics are consistent regardless of the material type and can be used to predict fatigue life with similar advantages.

The paper is organized as follows: Section 2 reviews current L-PBF fatigue performance research and identifies gaps. Section 3 describes the XCT scans and fatigue testing experiments. Section 4 details the MMTL framework for fatigue life prediction. Sections 5 and 6 validate MMTL through numerical simulation and a real-case study. Section 7 summarizes the findings and outlines future research directions.

2 Review of Related Research

2.1 The impact of process parameters on defects in L-PBF

The different types of defect generation in L-PBF Ti-6Al-4V specimens are impacted by the energy input controlled by process parameters such as laser power, scanning speed, hatch distance, and layer thickness (Criales et al. (2017)). Gong et al. (2014) categorized melting zones into "fully dense" (fewer GEPs), "over melting" (KHs), "incomplete melting" (LOFs), and "overheating" (failed builds) based on varying levels of energy input. Gordon et al. (2020) showed that KHs dominated the specimen even with high density in the excessive heat energy window (i.e., over melting). Pal et al. (2020) discovered that high scanning speeds (>500 mm/s) lead to LOFs due to incomplete melting, while lower speeds primarily result in KHs due to over melting. Moreover, other studies have explored and shown that the density of L-PBF 316L stainless steel specimens is significantly influenced by process parameters through experiments. With a fixed scanning speed of 300 mm/s, specimen

density decreased from 99.87% to 99.12% as laser power decreased from 380 W to 200 W, with an average of 1000 defects per mm^3 when laser power was below 260 W (Choo et al. (2019)). Additionally, specimens fabricated with low laser power (150 W), high scanning speed (781 mm/s), and a small hatch distance (0.08 mm) achieved a density of about 99.86% with the fewest defects observed in fractography images (Tucho et al. (2018)).

2.2 Defect classification via extracted features from XCT scans

XCT has been widely used to inspect specimens nondestructively, estimate part porosity in AM metal parts, and analyze defect features. Three defect features (volume/length, sphericity, and aspect ratio) from XCT scans are used to classify defects in L-PBF specimens with threshold values (Poudel et al. (2022)). Sphericity indicates how closely defects approximate a sphere, while aspect ratio indicates their flatness or elongation. Several studies indicate distinct characteristics for different defects in L-PBF parts when examined through high-resolution XCT scanning (0.3 or 1 μm voxel size) (Kasperovich et al. (2016); Poudel et al. (2022)). KHz typically exhibit lengths ranging from 30.3 to 65.8 μm , aspect ratios between 0.6 and 0.8, and sphericity values from 0.7 to 0.9. LOFs, on the other hand, tend to have lengths surpassing 100 μm , aspect ratios spanning 0.1 to 0.6, and sphericity ranging from 0.5 to 0.9. GEPs are characterized by lengths less than 50 μm , aspect ratios between 0.6 and 0.9, and sphericity values ranging from 0.8 to 1.0. Moreover, many studies have employed machine learning (ML) to classify defects (KHz, LOFs, and GEPs) in L-PBF specimens. Poudel et al. (2022) trained decision trees and neural networks to classify defects, and they achieved >98% and >99% classification accuracy, respectively. Ye et al. (2023) improved defect analysis by correlating features from low-resolution (5 μm voxel size) and high-resolution (1 μm voxel size) XCT scans, enhancing accuracy by 7.7% with k-nearest neighbor classification. In contrast, Snell et al. (2020) employed unsupervised k-means clustering to categorize defects based on XCT scan characteristics yet faced challenges with approximately 15% of defects labeled as "unclear" due to feature indistinctiveness.

2.3 Fatigue models for metal AM

Conventional and mechanical fatigue models Defect-based models have been developed to account for the effects of defect size and location on fatigue life (Hu et al. (2020))

due to the inevitable presence of defects and their detrimental impact on the fatigue life of metal AM parts. Murakami et al. (1986) developed a semi-empirical fracture mechanics-based model that considers the effects of defect size (i.e., Murakami size $\sqrt{area}$) and location on determining fatigue life. In this model, the fatigue limit and the stress intensity factor threshold were expressed as functions of the material’s hardness and the area of the surface defect. This model, with minor modifications, applies to other alloys like Ti-6Al-4V (Nakatani et al. (2019); Yamashita et al. (2017)) and enhances short crack growth predictions in the NASGRO equation (Maierhofer et al. (2014)). It has also been modified to predict the lower bound of fatigue limits Masuo et al. (2017) and estimate the effect of defects on fatigue life (Günther et al. (2017); Yamashita et al. (2017)).

Researchers also investigated fatigue performance based on the defect-tolerant design using the Kitagawa-Takahashi (KT) diagram, microstructure-based models, and multistage fatigue models. KT diagram was used to describe a smooth transition from short crack to long crack growth (El Haddad et al. (1979)), explore the relationship between defects and fatigue failure in L-PBF Ti-6Al-4V specimens (Komijani et al. (2019); Caton et al. (1999)), and identify the critical defect as well as estimate fatigue life based on its size with extreme value statistics (Niu et al. (2022)). Microstructure-based models address fatigue damage initiated from diverse microstructural features and propagated as small and long cracks throughout the specimens. McDowell et al. (2003) initially developed a microstructure-based model for high cycle fatigue of cast aluminum samples under multiaxial loading. According to the multistage fatigue (MSF) model (Xue et al. (2007)), fatigue life was divided into four stages: crack incubation, microstructurally small crack propagation, physically small crack propagation, and long crack propagation. Each term is governed by a specific formulation tailored for various materials, as described in (Xue et al. (2007)). Additionally, researchers attempted to predict the fatigue life using continuous damage mechanical models (Sandoval et al. (2020)) and stress-life curves (Tridello et al. (2021); Haridas et al. (2020)). Further details can be found in (Javidrad et al. (2024)).

Data-driven fatigue models Recently, researchers have employed ML models to evaluate the influence of process parameters, defect characteristics, and loading-related features on the fatigue performance of L-PBF specimens. Jia et al. (2023) proposed a deep belief

neural network to predict fatigue life using process parameters, powder size, and loading-related features, achieving an RMSE of 0.1. Bao et al. (2021) employed a support vector machine to predict fatigue life based on geometric features (e.g., size, sphericity) of critical defects and their location, achieving an MSE of 1.2736×10^{-3} . Salvati et al. (2022) developed a defect-based physics-informed neural network (PINN) incorporating Murakami’s $\sqrt{area}$ parameter for fracture mechanics, achieving an RMSE of 0.886. However, both conventional and data-driven models require destructive, data-intensive testing, which is time-consuming and expensive (Shi et al. (2023)). Limited data can cause the ML model to overfit during training, reducing its ability to accurately predict unseen fatigue life beyond the training set (Hornás et al. (2023)).

Transfer learning (TL), a supervised learning technique, can be employed to mitigate this issue by leveraging knowledge from a related task with ample data to address the challenge of limited data in a specific task (Senanayaka et al. (2023); Thrun and Pratt (1998)). Researchers have applied TL in various fatigue-related applications. Li et al. (2023) developed a transfer neural network to predict gear contact fatigue life. They pre-trained the model on abundant rolling contact fatigue data and transferred the learned information, achieving a MAPE of 0.4633. Wei et al. (2022) proposed a transfer long short-term memory network to predict the stress-life (S-N) curve of low alloy steels. Leveraging rotating bending S-N data, they trained the source model and predicted fatigue life based on limited data. Dong et al. (2023) introduced a TL approach to predict the remaining useful life of rolling bearings. Performing domain adaptation under similar operating conditions and fault behaviors, they achieved an RMSE of 0.094. Xiao et al. (2023) utilized TL to predict the fatigue life of corroded bimetallic steel bars. They pre-trained a neural network on metallic bars’ data and applied it to bimetallic steel bars, achieving an RMSE of 0.03. However, these studies can only use TL to address problems where the source task is similar to the target task. Additionally, TL alone cannot handle fatigue life prediction based on multiple input modalities, including process parameters, defect features, and fatigue-loading conditions.

2.4 Research gaps

After reviewing these related studies, the following research gaps are identified in assessing the fatigue performance of L-PBF parts:

(1) Some traditional methods only use L-PBF process conditions to classify defects and assess fatigue performance without identifying the critical defect. However, the co-occurrence of KHs, LOFs, or GEPs, even under the same process condition, makes these methods untenable for classifying defects and predicting fatigue life.

(2) Current ML models for defect classification and fatigue life prediction primarily rely on data-driven approaches, neglecting physics knowledge and fracture mechanics in defect generation and fatigue failure. Despite their acceptable prediction accuracy, they lack the structure to integrate physics knowledge to uncover insights into L-PBF fatigue performance, impeding widespread industry adoption.

(3) The time-consuming fatigue testing and the limited fatigue life data pose significant challenges in fatigue performance assessment for both traditional and ML models. Although XCT emerges as a nondestructive defect inspection tool to inspect many defects in L-PBF parts, its potential to identify critical defects and assess fatigue performance has not yet been explored.

This work addresses these gaps by proposing an MMTL framework that integrates L-PBF process conditions and XCT-derived defect information to understand defect features and types. It transfers learned feature embeddings to predict fatigue life nondestructively under varying loads with limited test specimens. Additionally, MMTL can adapt to other domains, incorporating physics knowledge and data from various sources to generate accurate and interpretable results with limited data.

3 Research Experiments and Problem Formulation

3.1 Experiments

Two sets of L-PBF Ti-6Al-4V Grade 5 parts were fabricated by an EOS M290 machine by changing laser power and scanning speed from EOS recommended infill process parameters shown in Table 1 to induce volumetric defects (i.e., KHs, LOFs, and GEPs). The first

set (i.e., Set 1) investigated the characteristics of different volumetric defects scanned from XCT. The second set (i.e., Set 2) identified the critical defects in volumetric defects based on their locations, size, and morphology features extracted from XCT, as well as the fatigue testing under different stress amplitudes and fractography.

Table 1: Two sets of L-PBF parts were fabricated under different process conditions.

Part set	Process parameters	Process conditions					Units
		P1	P2	P3	P4	P5	
Set 1	Laser power	224	252	280	280	336	W
	Scanning speed	1300	1560	1300	1560	780	mm/s
Set 2	Laser power	224	252	280	336	364	W
	Scanning speed	1200	1200	1200	840	960	mm/s

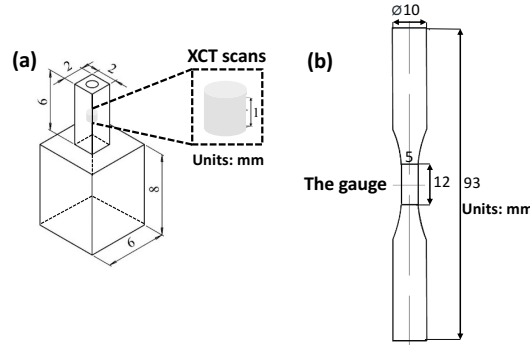


Figure 2: (a) The design geometry for the XCT scanning portion for Set 1, and (b) the XCT scanning portion at the gauge and during fatigue testing.

In Set 1, five parts were selected to be scanned by XCT. The cylindrical portions of the fabricated parts were machined into rectangular bars of 2 mm thickness (Figure 2 (a)) to permit high-resolution XCT scans (a ZEISS Xradia 620 Versa machine with $1 \mu\text{m}$ voxel size) in the infill region. Following the completion of the scan, the volumetric tomography data underwent reconstruction using the ZEISS Reconstruction software. Subsequently, the volumetric defects were labeled to three defect types - KHs, LOFs, and GEPs by five domain experts in L-PBF and materials (Poudel et al. (2022)). Only defects with consensus from at least four out of five expert evaluators were included to ensure reliability, while those without were excluded. Out of 2156 identified defects (only those larger than $10 \mu\text{m}$ were considered), 1531 were conclusively labeled: 68 as KHs, 1308 as LOFs, and 155 as GEPs. Notably, parts fabricated under process condition P3 exhibited the fewest defects while overheating (P5) led to more KHs than LOFs and GEPs, and underheating (P1, P2, P4)

resulted in more LOFs than KHs and GEPs.

In characterizing the reconstructed volumetric defects obtained from XCT scans, we employed a computer vision approach, content-based image retrieval (Schneider et al. (2012)), to extract diverse volumetric defect size- and morphology-related features by several shape descriptors. Table 2 summarizes four size-related and four morphology-related features, describing volumetric defect characteristics from different perspectives. These features encompass more information on volumetric defects than the presently used features (such as length, sphericity, aspect ratio, and others, as discussed in Section 2). They provide insights into the types of volumetric defects and contribute to the discriminative features essential for individual volumetric defect classification. In cases where distinct types of volumetric defects exhibit similar features (e.g., small KHs and GEPs), the inclusion of energy input (L-PBF process parameters) as prior information (refer to Table 1) in volumetric defect classification has the potential to enhance the accuracy of classifications.

Table 2: The definitions and interpretations of size- and morphology-related features of volumetric defects from XCT scans of L-PBF parts. (b and c indicate minor axis length and the height of a defect, respectively).

Category	Features	Symbols	Definitions
Size	Volume	V_o	The size of the defect in L-PBF parts. It is crucial in determining the part's mechanical properties, such as fatigue life
	Surface area	S	The total area of the exposed surface of the defect in L-PBF parts. It can influence how the defect interacts with the surrounding material, potentially affecting the part's strength, fatigue life, and overall performance
	Major axis length	a	The longest distance between two opposite points on the shape along an axis; The larger the value, the easier it is to cause cracks
	Ellipsoid volume	V_e	The volume of a defect approximated by the shape of an ellipsoid. It is important to evaluate defects' impact on the mechanical properties of the L-PBF part
Morphology	Aspect ratio	$r = \frac{b}{a}$	The ratio between the minor axis length and major axis length of the defect; the low aspect ratio indicates its elongated shape, which can easily cause high-stress concentration
	Sphericity	$(36\pi V_o^2/S^3)^{\frac{1}{3}}$	The degree to which a shape approximates a sphere; a defect with a perfectly spherical shape has a sphericity of 1, which is the least likely to initiate a crack
	Sparseness	$\frac{V_e}{V_o}$	The measure of how much the defect fills the ellipsoid that approximates its shape. A sparseness value greater than 1 indicates that the defect occupies less space than the ellipsoid, suggesting a more irregular and sparse defect
	Flatness	$\frac{b}{c}$	The ratio between the minor axis length and the height of a defect; a defect with high flatness is more likely to initiate a crack with higher stress concentration

In Set 2, thirteen as-built parts undergo stress relief at 705°C for 1 hour, with a heating rate of 5°C/minute in an electric furnace, followed by furnace cooling before removal from the build plate. Subsequently, these parts underwent the same XCT scans as Set 1 and were then machined into round fatigue testing specimens (standard ASTM E466 on Mechanical Testing

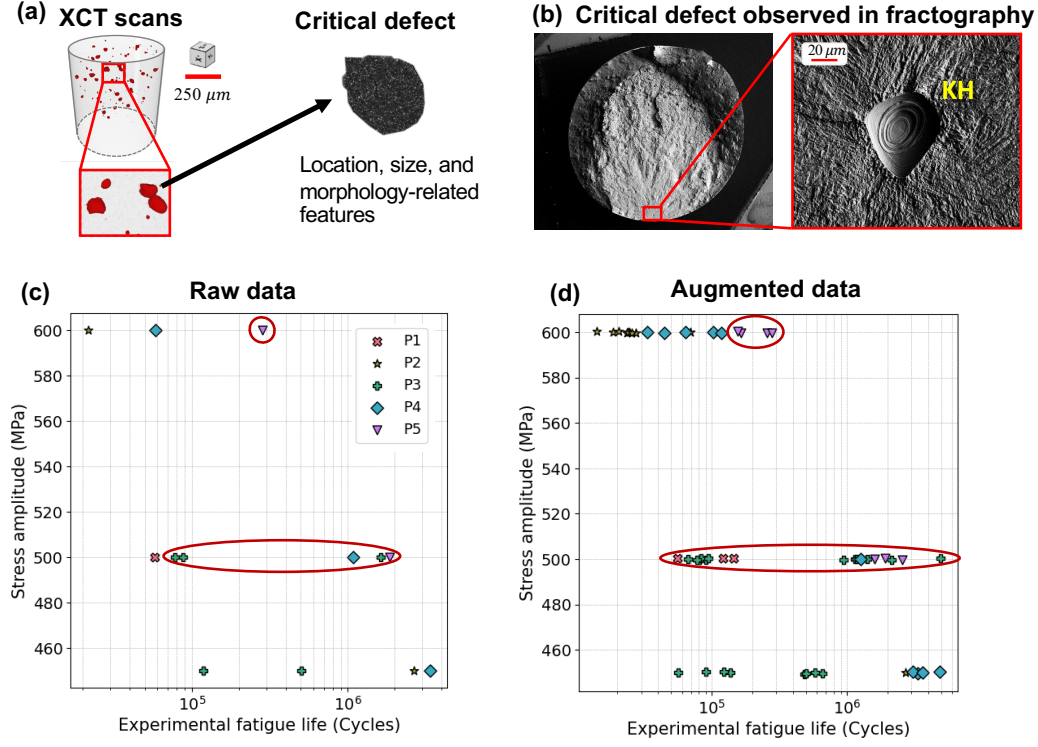


Figure 3: (a) The critical defect in XCT scans, (b) critical defect observed in fractography, (c) the raw stress-fatigue life (S-N) curve for L-PBF Ti-6Al-4V specimens (Set 2), fabricated under five distinct process conditions shown in Table 1, and (d) the augmented data (50 data points) from raw data (13 data points) in Set 2. In (c), under 500 MPa, data points for conditions P3 in a circle are very different (large scatter). Certain process conditions, such as P5 in a circle, consistently demonstrate higher fatigue life across different stress levels. The augmented data maintain a similar scatter in fatigue performance, as depicted by the ellipse in (d).

(2016)), as shown in Figure 2 (b). We conducted destructive fatigue testing for Set 2, employing a fully-reversed ($R = -1$) stress-controlled mode, with stress amplitude set at 450, 500, and 600 MPa to investigate specimens' short and long-cycle fatigue regimes. The critical defect, which caused the crack initiation and part fracture, was examined by fractography (Figure 3 (b)) and then identified in the XCT scans (Figure 3 (a)) by a cross-dimensional defect matching process method (Ye et al. (2023)) based on its location, size, and morphology. These features of the critical defects (e.g., length $> 200 \mu\text{m}$, aspect ratio < 0.3) are used to identify the potential critical defects in practical applications. Jia et al. (2023) has shown this method can achieve 81.25% accuracy in correctly identifying critical defects from XCT scans. We predicted fatigue life with all candidates of the critical defect

and generated a distribution of fatigue life for the particular part to achieve reliable adoption of an L-PBF part.

S-N curve in Figure 3 (c) is used to show the challenges in investigating the fatigue life of L-PBF parts using traditional tests and curves: (1) the large scatter of the fatigue life of L-PBF parts does not reveal clear patterns for accurate fatigue life prediction; (2) the limited fatigue data is the unfortunate reality in investigating the fatigue performance of L-PBF parts with extremely time-consuming fatigue tests. Under 500 MPa, data points for conditions P3 are very different (large scatter). Their overlap or approximation with points for other conditions implies some other factors play larger impacts on fatigue life, which are determined as critical defects from fractography in Figure 3 (b). Certain process conditions, such as P5, consistently demonstrate higher fatigue life across different stress levels.

Based on these insights and observations, our proposed MMTL framework designs a hierarchical structure incorporating defect features and process parameters into fatigue life prediction. To enhance the robustness of our results, we employed data augmentation techniques in Set 2 by adding Gaussian noise with a mean of zero and a standard deviation based on the features (excluding process parameters and stress amplitude). Specifically, we increased the sample size of Set 2 from 13 to 63 data points, comprising 50 augmented data points and 13 raw data points. This approach maintains the original data’s characteristics while providing a larger dataset for analysis, as shown in Figure 3 (c) and (d).

3.2 Problem Definition

Inspired by the insights gained from the aforementioned experiments, we formulate the fatigue life prediction problem by accounting for the influence of both process parameters in generating volumetric defects and their size- and morphology-related features, along with the subsequent impact of critical defects on fatigue performance. We address this problem within our MMTL framework, as illustrated in Figure 4. The input data consists of L-PBF process parameter tabular data and XCT volumetric defect image data, incorporating both modalities into the framework. It integrates informative feature embedding derived from these two modalities process parameters and defect features associated with various volumetric defects (the source task S) using Set 1’s extensive defect data (refer to Table 1). This enrichment of knowledge about critical defects subsequently enhances the accuracy of

fatigue life prediction (the target task T) utilizing the augmented dataset of Set 2.

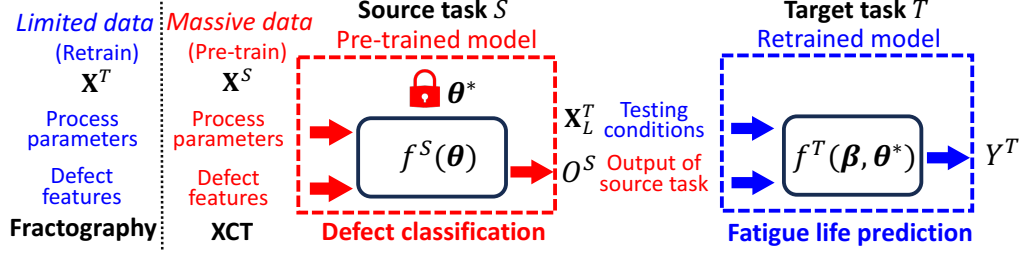


Figure 4: The formulation of the MMTL framework. A pre-trained model with massive XCT data for source task S defect classification and a retrained model with limited fractography data for target task T fatigue life prediction.

Problem formulation We formulate the fatigue life prediction problem in the MMTL framework as

$$\begin{aligned}
 Y^T &= f^T(\beta; \mathbf{X}_L^T, f^S(\theta^*; \mathbf{X}^T)) \\
 &= \underbrace{(f^T(\beta; \mathbf{X}_L^T))}_{\text{Retrained term}} \circ \underbrace{(f^S(\theta^*; \mathbf{X}^T))}_{\text{Pre-trained term}}
 \end{aligned} \tag{1}$$

where Y^T is the prediction of the target task T from the MMTL framework; $\mathbf{X}^T = [\mathbf{X}_1^T, \dots, \mathbf{X}_M^T]^T$ is the data with M modalities in the target task T (e.g., L-PBF process conditions and defects), and $\mathbf{X}_i^T = [X_{i,1}^T, \dots, X_{i,d}^T]^T$ denotes the i -th modality with d features; $\mathbf{X}_L^T$ is additional input in the target task T (e.g., fatigue test conditions); $f^T(\beta)$ is the TL function with the pre-trained model $f^S(\theta)$ with optimal parameters θ^* and leverages the information captured by the pre-trained model structure; $\circ$ is the composition operator defined as $(f \circ g)(x) = f(g(x))$.

Parameter estimation We utilize a bi-level optimization (i.e., an optimization problem nested within another, where the solution of the inner problem affects the outer problem) for the source task S and target task T to estimate θ and β since the models used for defect classification (source task S) and fatigue life prediction (target task T) cannot be trained and optimized simultaneously, as the fatigue life values are only available for the data points (13 fatigue specimens and 50 augmented data) in the target task T and not in the source

task S . The optimization problem can be represented as

$$\begin{aligned} \min_{\beta} \mathbb{E}[\mathcal{L}^T(Y^T, (f^T(\beta; \mathbf{X}_L^T) \circ f^S)(\theta^*; \mathbf{X}^T))] \\ \text{s.t. } \theta^* \in \arg \min_{\theta} \mathbb{E}[\mathcal{L}^S(O^S, f^S(\theta; \mathbf{X}^S))] \end{aligned} \quad (2)$$

where $\mathcal{L}^T$ and $\mathcal{L}^S$ are the loss functions for target task T and source task S , respectively. O^S is the output in the source task S (i.e., defect classification). $\mathbf{X}^S = [\mathbf{X}_1^S, \dots, \mathbf{X}_M^S]^T$ is the data for the source task S .

In this paper, we consider two modalities ($M = 2$) for the MMTL framework, i.e., L-PBF process parameters and defect features from XCT scans. In the training and evaluation steps, cross-entropy (CE) is used in the source task as $\mathcal{L}^S$ for defect classification and mean-square error (MSE) in the target task as $\mathcal{L}^T$ for fatigue life prediction.

4 Methodology of Multimodal Transfer Learning

4.1 Hierarchical graph convolutional network in source task

We propose HGCN for the source task—defect classification in L-PBF parts, which integrates the understanding of defect characteristics into defect classification and fatigue life prediction. It can capture this relationship by using process parameters and defect features as inputs, providing a comprehensive feature embedding across different hierarchies and potentially improving classification performance. Furthermore, the extracted informative feature embedding from HGCN can enhance fatigue life prediction performance (detailed in Section 4.2). In HGCN, defects are represented as graphs with nodes and edges, incorporating information from process parameters, defect features, and their similarity. HGCN is structured hierarchically with two GCN modules and will train them concurrently, as shown in Figure 5. GCN₁: $g_1(\theta_1)$ on Hierarchy 1 embeds process parameters to a latent space (i.e., defect feature space), while GCN₂: $g_2(\theta_2)$ on Hierarchy 2 focuses on learning defect features for defect classification.

Graph representation of defects The data from all volumetric defects is organized into graphs, where nodes represent volumetric defects and edges represent node similarity (Hang

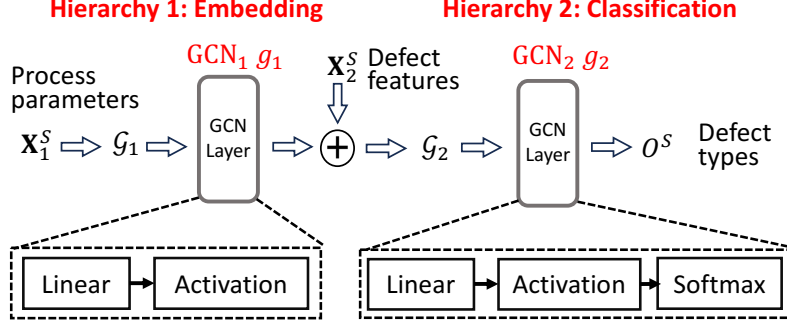


Figure 5: The proposed HGCN is designed for volumetric defect classification with two hierarchies. Hierarchy 1 encodes the first modality information in the graph-structured data, and Hierarchy 2 integrates the second modality and output from Hierarchy 1 to classify defects.

and Singh (2010)). In the HGCN model, two undirected graphs, $\mathcal{G}_1$ and $\mathcal{G}_2$, are constructed for the two hierarchies, respectively.

In the graph $\mathcal{G}_1 = (\mathcal{V}_1, \mathcal{E}_1)$, each node $u_i \in \mathcal{V}_1$ (for $i = 1, \dots, N$) represents the process parameters of individual defects. An edge $e_{ii'} \in \mathcal{E}_1$ connecting two nodes, u_i and $u_{i'}$, signifies the similarity between the process parameters of two defects. $\mathcal{G}_1$ encapsulates the prior knowledge of defect types based on process parameters, indicating that two volumetric defects are of the same type if generated under similar process parameters. Within graph $\mathcal{G}_1$, an adjacency matrix $\mathbf{A}_1 \in \mathbb{R}^{N \times N}$ is constructed using normalized Euclidean distances among nodes, restricted to the range $[0, 1]$. Distances below a threshold $t \in [0, 1]$ lead to connections between similar nodes (i.e., $\mathbf{A}_1[u_i, u_{i'}] = 1$ if the normalized Euclidean distance is less than the threshold). Edges in $\mathcal{E}_1$ are derived from the adjacency matrix $\mathbf{A}_1$ (i.e., $e_{ii'} \in \mathcal{E}_1$ if $\mathbf{A}_1[u_i, u_{i'}] = 1$), and an adjacency matrix with self-loop nodes $\mathbf{A}_1^* \in \mathbb{R}^{N \times N}$ introduces connections to each node itself (i.e., $\mathbf{A}_1^*[u_i, u_{i'}] = 1$).

In the graph $\mathcal{G}_2 = (\mathcal{V}_2, \mathcal{E}_2)$, each node $v_j \in \mathcal{V}_2$ (for $j = 1, \dots, N$) represents an individual defect j with the attribute of extracted defect features. An edge $e_{jj'} \in \mathcal{E}_2$ connecting two nodes, v_j and $v_{j'}$, signifies the similarity between the corresponding volumetric defects, considering both defect features and process parameters.

Graph convolutional network and hierarchical structure of HGCN GCN operates on graph-structured data (Yao et al. (2019); Song et al. (2019)) and employs graph convolutions to learn informative features by aggregating information from neighbors (i.e.,

connected nodes), especially for nodes with multiple neighbors from the same class, which can achieve more accurate defect classification compared to the neural network. We present the comparable results from our simulation in Section 5, which utilized two overlapping components generated from a bivariate Gaussian mixture model. Besides, GCN leverages graph representation of data and provides better modeling and interpretation of defect classification.

A GCN unit comprises multiple graph convolutional layers, and the function of the k -th layer ($k = 1, \dots, K$) can be represented as Kipf and Welling (2016),

$$\mathbf{H}^{(k)} = \text{ReLU}(\tilde{\mathbf{A}}^* \mathbf{H}^{(k-1)} \boldsymbol{\theta}^{(k-1)}) \quad (3)$$

where $\mathbf{H}^{(k)} = [h_1^{(k)}, \dots, h_N^{(k)}]^T$ is the output of the k -th layer of GCN, $\tilde{\mathbf{A}}^*$ is the normalized adjacency matrix as $\tilde{\mathbf{A}}^* = \mathbf{D}^{-\frac{1}{2}} \mathbf{A}^* \mathbf{D}^{-\frac{1}{2}}$, and $\mathbf{D}$ is a diagonal degree matrix give as $\mathbf{D} = \text{diag}(d_1, \dots, d_N)$ with $d_i = \sum_{v \in \mathcal{V}} \mathbf{A}^*[v_i, v_i]$. $\boldsymbol{\theta}^{(k)}$ is the weight matrix of the k -th layer.

Hierarchy 1: Embedding GCN₁ within Hierarchy 1 is tailored to capture the influences of process parameters on defect features and their corresponding types. This is achieved by embedding process parameters into the defect feature space, as represented by Eq.(3),

$$g_1(\boldsymbol{\theta}_1; \mathbf{X}_1^S) : \mathbf{H}_1^{(j)} = \text{ReLU}(\tilde{\mathbf{A}}_1^* \mathbf{H}_1^{(j-1)} \boldsymbol{\theta}_1^{(j-1)}), j = 1, \dots, J \quad (4)$$

where $\tilde{\mathbf{A}}_1^* \in \mathbb{R}^{N \times N}$ is the normalized adjacency matrix, $\mathbf{H}_1^{(j)} \in \mathbb{R}^{N \times q}$ is the output of $(j - 1)$ -th GCN layer and $\mathbf{H}_1^{(0)} = \mathbf{X}_1^S$, $\boldsymbol{\theta}_1^{(j-1)} \in \mathbb{R}^{d \times q}$ is the weight matrix.

Hierarchy 2: Classification GCN₂ within Hierarchy 2 is crafted to model the classification function g_2 by incorporating both defect features $\mathbf{X}_2^S$ and process parameters $\mathbf{X}_1^S$ as

$$\begin{aligned} \mathbf{H}_2^{(k)} &= \text{ReLU}(\tilde{\mathbf{A}}_2^* \mathbf{H}_2^{(k-1)} \boldsymbol{\theta}_2^{(k-1)}), k = 1, \dots, K \\ g_2(\boldsymbol{\theta}_2; \mathbf{X}_2^S) : O^S &= \text{Softmax} \circ \mathbf{H}_2^{(K)} \end{aligned} \quad (5)$$

where $\tilde{\mathbf{A}}_2^* \in \mathbb{R}^{N \times N}$ is the normalized adjacency matrix for volumetric defects with both defect features and process parameters, $\mathbf{H}_2^{(k)} \in \mathbb{R}^{N \times p}$ is the output of $(k - 1)$ -th GCN

layer and $\mathbf{H}_2^{(0)} = [\mathbf{X}_2^S | \mathbf{H}_1^{(J)}]$, and the operator $[\cdot | \cdot]$ is the horizontal concatenation of two matrices. It is noted that weight matrix $\boldsymbol{\theta}_1^{(j-1)}$ follows $\mathbb{R}^{(m+q) \times p}$ since we concatenate volumetric defect features $\mathbf{X}_2^S$ (i.e., m dimensions) and the output of Hierarchy 1 $\mathbf{H}_1^{(J)}$ (i.e., q dimensions).

To determine the optimal $\boldsymbol{\theta}_1$ in Eq.(4) and $\boldsymbol{\theta}_2$ in Eq.(5), we employ CE for the volumetric defect classification source task problem, as discussed in Section 3.2 Eq.(2),

$$\begin{aligned} \boldsymbol{\theta}^* &\in \arg \min_{\boldsymbol{\theta}=[\boldsymbol{\theta}_1, \boldsymbol{\theta}_2]} \mathbb{E}[-\sum_{c=1}^C O^S \log \hat{O}^S] \\ &= \arg \min_{\boldsymbol{\theta}=[\boldsymbol{\theta}_1, \boldsymbol{\theta}_2]} \mathbb{E}[-\sum_{c=1}^C O^S \log (g_2(\boldsymbol{\theta}_2; \mathbf{X}_2^S) \circ g_1(\boldsymbol{\theta}_2; \mathbf{X}_1^S))] \end{aligned} \quad (6)$$

In this paper, we opt for the Adam optimizer (Kingma and Ba (2014)), a widely utilized optimization algorithm in deep learning, to update the parameters $\boldsymbol{\theta}$. The derivation of the parameter update for Eq.(6) can be found in Appendix 8.1.

4.2 Neural network layer in target task

Utilizing the pre-trained HGCN for the source task of defect classification, we proceed to retrain neural network layers for the fatigue life prediction of L-PBF specimens. This re-training incorporates process parameters $\mathbf{X}_1^T$, critical defect features $\mathbf{X}_2^T$, other inputs $\mathbf{X}_L^T$ (e.g., defect locations, and fatigue testing stress amplitude) as input data and formulated as follows (similar to Eq.(1) in Section 3.2),

$$Y^T = f^T(\boldsymbol{\beta}; \mathbf{X}_L^T, f^S(\boldsymbol{\theta}^*; \mathbf{X}_1^T, \mathbf{X}_2^T)) = \sigma(\mathbf{Z}^{(l)} \boldsymbol{\beta}^{(l)}) \quad (7)$$

where σ is the activation function, $\mathbf{Z}^{(l)}$ is the output of l -th layer and the first layer input $\mathbf{Z}^{(0)} = [\mathbf{X}_L^T, f^S(\boldsymbol{\theta}^*; \mathbf{X}_1^T, \mathbf{X}_2^T)] = [\mathbf{X}_L^T, \mathbf{H}_2^{(K)*}]$. $\boldsymbol{\beta}^{(l)}$ follows $\mathbb{R}^{(n+C) \times r}$ since we concatenate fatigue-loading related features $\mathbf{X}_L^T$ (i.e., n dimensions) and output (last layer of GCN₂) of volumetric defect classification in source task $f^S(\boldsymbol{\theta}^*; \mathbf{X}_1^T, \mathbf{X}_2^T)$ (i.e., C dimensions).

To determine the optimal $\boldsymbol{\beta}$ in Eq.(7), we select MSE for the fatigue life prediction target

task problem as mentioned in Section 3.2 Eq.(2),

$$\begin{aligned} & \min_{\beta} \mathbb{E}[\mathcal{L}^T(Y^T, (f^T(\beta; \mathbf{X}_L^T) \circ f^S)(\theta^*; \mathbf{X}_1^T, \mathbf{X}_2^T))] \\ & = \min_{\beta} \mathbb{E}[(\hat{Y}^T - Y^T)(\hat{Y}^T - Y^T)^T] \end{aligned} \quad (8)$$

Here, we also use Adam optimizer (Kingma and Ba (2014)) to update the parameters β . The derivation can be found in Appendix 8.1.

5 Numerical Simulation

Synthetic data generation To validate the proposed MMTL framework for L-PBF fatigue life prediction, we leverage two Gaussian mixture models (GMMs) with two components $\mathbf{G}_m \sim \sum_{k=1}^2 \pi_{m,k} \mathcal{N}(\mu_k, \Sigma_k)$, *s.t.*, $\sum_{k=1}^2 \pi_{m,k} = 1$ ($m = 1, 2$) to simulate data representing process-defect-fatigue relationships. The mixture weights of two GMMs ($m = 2$) represent two process conditions, $\pi_1 = [\pi_{1,1}, \pi_{1,2}] = [0.9, 0.1]$ and $\pi_2 = [\pi_{2,1}, \pi_{2,2}] = [0.8, 0.2]$. The two components ($k = 2$) in GMMs represent two main types of critical defects co-existing in each process condition, and we define $\mu_1 = [0, 0]$, $\mu_2 = [1, 1]$, and $\Sigma_1 = \Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. A total of 300 synthetic defect data points are generated from the two GMMs, half from $\mathbf{G}_1$, and the other half from $\mathbf{G}_2$.

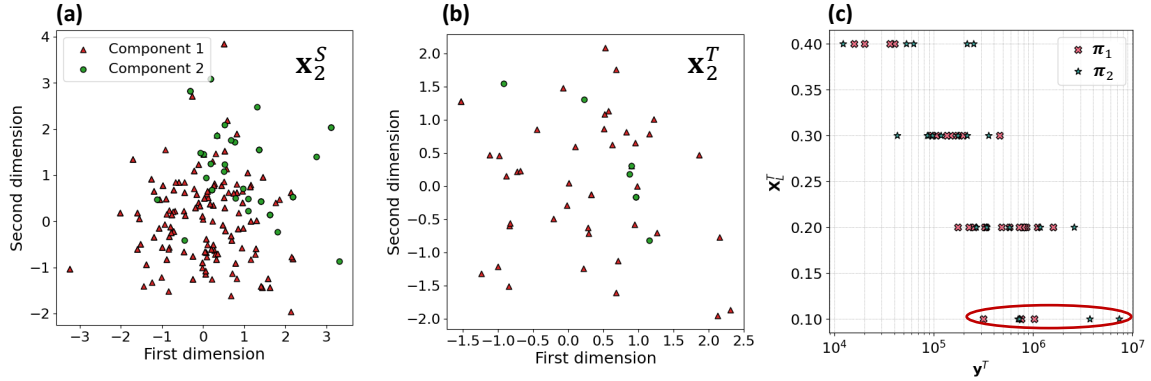


Figure 6: The visualization of source data (a) and target data (b) and simulated S-N curve (c). The simulated response y^T in (c) shows a significant scatter under synthetic stress amplitude X_L^T (e.g., points spots in the circle), which is similar to fatigue life scatter under the same stress amplitude mentioned in Figure 3.

We randomly select 250 data points as source data, representing their mixture weights as

$\mathbf{X}_1^S$ (i.e., process parameters) and their feature values as $\mathbf{X}_2^S$ (i.e., defect features) in Figure 6 (a) to validate the source task in the MMTL framework for extracting feature embedding from process and defect and classifying defects. We select the rest of 50 data points as target data (i.e., critical defects), representing their mixture weights as $\mathbf{X}_1^T$ (i.e., process parameters) and their feature values as $\mathbf{X}_2^T$ (i.e., defect features) in Figure 6 (b), and the target data has the corresponding fatigue life, simulated from a generalized additive model $\log(\mathbf{y}^T) = 1 + \sin(2\pi\mathbf{X}_1^T) + \cos(2\pi\mathbf{X}_2^T) - 10\mathbf{X}_L^T$ to represent the long-tail effects in fatigue life controlled by synthetic stress amplitude $\mathbf{X}_L^T \sim \text{U}(0, 0.5)$, as shown in Figure 6 (c). These 50 target data are used in the target task to validate the efficiency of the MMTL for prediction.

Graph representation of synthetic data In the MMTL framework, $\mathcal{G}_1^S$ and $\mathcal{G}_2^S$ are constructed from the 250 source data, represented by $\mathbf{X}_1^S$ and $\mathbf{X}_2^S$, for the two GCNs on the two hierarchies of HGCN in the source task; $\mathcal{G}_1^T$ and $\mathcal{G}_2^T$ are constructed from the 50 target data, represented by $\mathbf{X}_1^T$ and $\mathbf{X}_2^T$, in the target task. To analyze the effects of feature values and mixture weights on their components, we construct two graphs in the source task, as shown in Figure 7. The reference graph, denoted as $\mathcal{G}_r^S$, is constructed by including only feature values $\mathbf{X}_2^S$. It assumes that all nodes from the same component are connected, while 50% of the nodes from different components are randomly connected to simulate a challenging classification scenario where some nodes from different components still have similar feature values. Moreover, the graph $\mathcal{G}_2^S$ is built from $\mathcal{G}_r^S$ by incorporating mixture weights and using prior class information to simplify node connections. It is noted that $\mathcal{G}_2^S$ integrating feature values and mixture weights can classify nodes from different components more easily than $\mathcal{G}_r^S$. Such integration simplifies the graph structure (i.e., lower average node degree), as shown in Table 3, potentially improving classification and prediction accuracy.

Table 3: The summary of graphs for synthetic source and target data. $\mathcal{G}_2^S$ exhibits a simpler graph structure with a lower average node degree.

Graph	Number of nodes	Number of edges	Number of features per node	Average node degree
$\mathcal{G}_1^S$	250	23421	2	93.68
$\mathcal{G}_2^S$	250	12065	4	48.26

The evaluation of GMM component classification in MMTL source task We compare HGCN with Hierarchical Neural Network (HNN), Graph Convolutional Network

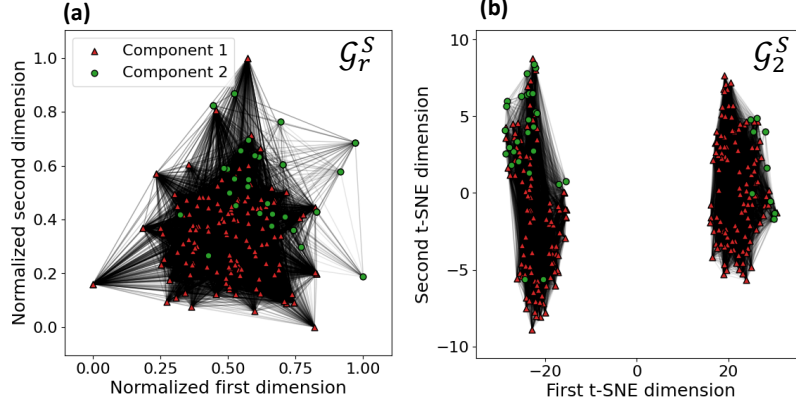


Figure 7: The t-distributed Stochastic Neighbor Embedding (t-SNE) visualization of graph representations illustrates graphs (a) $\mathcal{G}_r^S$ and (b) $\mathcal{G}_2^S$ for a synthetic source dataset with 250 data points. The black lines denote the edges between nodes in these graphs. $\mathcal{G}_2^S$ aims to demonstrate that by incorporating feature values and mixture weights, the graph structure can be simplified with fewer edges and a simple graph structure. Consequently, it becomes easier to differentiate data points compared to $\mathcal{G}_r^S$.

(GCN), and Neural Network (NN) for GMM component classification, based on F1-score (full details of model structures are provided in Appendix 8.2.1). The graphs based on source data $\mathbf{X}_1^S$ and $\mathbf{X}_2^S$ are utilized for training and validating HGCN for classification. Then, the trained HGCN is tested by using the graphs from the target data $\mathbf{X}_1^T$ and $\mathbf{X}_2^T$. In Table 4, HGCN achieves an F1-score of 0.9811 on the test set of the source data, demonstrating comparable classification performance to benchmark models HNN and NN. This underscores the significant improvement in GMM component classification by aggregating node information from neighbors in graphs. Furthermore, HGCN surpasses GCN, indicating its ability to derive more informative embeddings from mixture weights and feature values, which is particularly advantageous for prediction in the MMTL target task. Additionally, HGCN and GCN achieve a superior F1-score of 1.0000 on the target dataset, outperforming other benchmark models. This underscores their capability to effectively aggregate useful information from neighboring nodes for accurate classification.

Table 4: Comparison of F1-score classification performance for HGCN, HNN, GCN, and NN on simulated source and target data. Note: Standard deviations from 5-fold cross-validation are presented in parentheses.

	Our model	Benchmark models		
	HGCN	HNN	GCN	NN
Source data	0.9811 (0.0379)	0.6565 (0.1112)	0.9298 (0.0784)	0.6240 (0.0762)
Target data	1.0000	0.6250	1.0000	0.5127

The evaluation of prediction in MMTL target task We compare MMTL with pre-trained HNN, pre-trained GCN, pre-trained NN with additional NN layers, and a baseline simple NN for prediction of $\mathbf{y}^T$ based on root mean squared log error (RMSLE), and mean absolute percentage log error (MAPLE). RMSLE and MAPLE are used in our paper to reduce the impact of large errors, as the target variable, spans multiple orders of magnitude. The proposed MMTL achieves the lowest RMSLE at 0.3236 with 30.03%, and MAPLE at 0.0442 with 30.83% improvement compared to the baseline model NN, as shown in Table 5. This demonstrates that the learned embeddings from mixture weights and feature values in the pre-trained HGCN were successfully transferred and utilized to enhance prediction performance with limited data.

Table 5: Comparison of prediction performance for MMTL against benchmark models on the simulated target data. Standard deviations for 5-fold cross-validation are presented in parentheses.

		Our model	Benchmark (Pre-trained model + NN layer)			Baseline
		MMTL	HNN+NN layer	GCN+NN layer	NN+NN layer	NN
RMSLE	Test	0.3236 (0.0718)	0.3704 (0.0854)	0.3552 (0.1211)	0.3712 (0.0785)	0.4625 (0.1327)
	Improved %	↑ 30.03	↑ 19.91	↑ 23.20	↑ 19.74	-
MAPLE	Test	0.0442 (0.0103)	0.0569 (0.0155)	0.0527 (0.0194)	0.0580 (0.0132)	0.0639 (0.0231)
	Improved %	↑ 30.83	↑ 10.95	↑ 17.53	↑ 9.23	-

6 Case Study of Fatigue Life Prediction of L-PBF Parts

Defect data description We assess the effectiveness of the proposed MMTL in predicting the fatigue life of L-PBF specimens. The detail of the experiments is described in Section 3.1. We refer to Set 1 with 1531 volumetric defects as the source data (Figure 8 (a)) to train HGCN in MMTL and augmented data with 63 critical defects as the target data (Figure 8 (b)) to train NN layers for fatigue life prediction.

Graph representation of defect data In the proposed MMTL model for the case study, $\mathcal{G}_1^S$ and $\mathcal{G}_2^S$ are constructed from process parameters $\mathbf{X}_1^S$ and defect features $\mathbf{X}_2^S$ of Set 1 for the two GCNs on the two hierarchies of HGCN in the source task; $\mathcal{G}_1^T$ and $\mathcal{G}_2^T$ are constructed from process parameters $\mathbf{X}_1^T$ and defect features $\mathbf{X}_2^T$ of augmented data for fatigue life prediction in the target task. Edges are determined by thresholding ($t = 0.1$) similarity distances between nodes based on the grid search from 0.1-1 with step

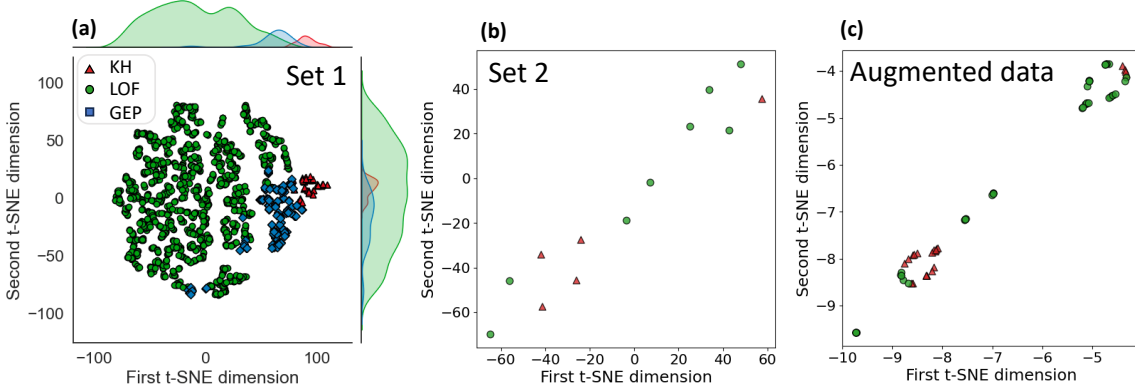


Figure 8: The visualization of different types of defects using t-SNE for (a) source data (i.e., Set 1), (b) Set 2, and (c) target data (i.e., augmented data) based on eight defect features mentioned in Table 2. The three kernel-estimated distributions on each side in (a) represent the marginal distributions of the three classes along their respective dimensions. KHs and LOFs are more likely to be critical defects affecting the fatigue performance of L-PBF parts, as shown in (b).

0.1. To explore the impact of process parameters and defect features on defect types, we also construct a reference graph $\mathcal{G}_r^S$ only with $\mathbf{X}_2^S$ from Set 1 for defect classification.

Table 6: The summary of graphs for the source dataset. $\mathcal{G}_2^S$ has simpler graph structures with more concentrated information.

Graph	Number of nodes	Number of edges	Number of features per node	Average node degree
$\mathcal{G}_r^S$	1531	221694	6	145
$\mathcal{G}_2^S$	1531	157912	8	103

For defect classification, it is observed from Figure 9 (b) that by considering the impact of process parameters on defect features, we can infer the type of defects more easily, as shown in $\mathcal{G}_2^S$. It verifies the assumption that different types of defects are generated due to energy input with their own size and morphology mentioned in Section 1 from the graph perspective. Moreover, such integration simplifies graph structure with more concentrated information (i.e., higher average node degree), as shown in Table 6, distinguishes defects with unique sizes or morphology, and improves classification and prediction accuracy.

The evaluation of defect classification in MMTL source task We compare HGCN with HNN, GCN, and NN based on F1-score (full details of model structures in Appendix 8.2.2), as shown in Table 7. The Set 1 is utilized for training and testing HGCN. Then, the trained HGCN is examined using the augmented data. The different types of defects

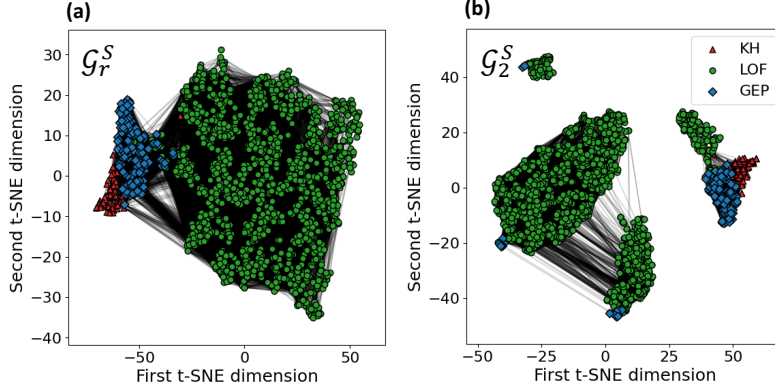


Figure 9: The t-SNE graph representation illustrates the graphs (a) $\mathcal{G}_r^S$, (b) $\mathcal{G}_2^S$ for the source dataset. The black lines in these graphs represent the edges between nodes. $\mathcal{G}_2^S$ with both process parameters and defect features has a simpler graph structure and can differentiate defects more easily than $\mathcal{G}_r^S$ with only the defect features.

in Set 1 are differentiable, so all the methods can achieve a high F1-score in classification. Moreover, HGCN achieves a superior F1-score of 0.8288 on augmented data because process parameters provide supplementary information for defect classification with limited data. It indicates that HGCN can capture this relationship by using process parameters and defect features as inputs, providing a comprehensive feature representation across different hierarchies and potentially improving classification performance and interpretability. This knowledge (i.e., informative feature embedding learned from process parameters and defect features) can be transferred and leveraged in subsequent NN training for more accurate fatigue life prediction.

Table 7: Comparison of F1-score classification performance for HGCN against HNN, GCN, and NN on Set 1 and augmented data. Note: Standard deviations for 5-fold cross-validation are presented in parentheses.

	Our model	Benchmark		
	HGCN	HNN	GCN	NN
Set 1	0.9593 (0.0138)	0.9552 (0.0166)	0.9497 (0.0185)	0.9617 (0.0243)
Augmented data	0.8288	0.8253	0.3942	0.4749

The evaluation of MMTL for fatigue life prediction of L-PBF specimens

We compare our proposed MMTL framework with pre-trained HNN, pre-trained GCN, pre-trained NN with an additional NN, and a baseline NN for fatigue life prediction based on RMSLE and MAPLE. As mentioned in Section 5, RMSLE and MAPLE are used in

our paper to mitigate the impact of large errors, as the target variable, fatigue life, spans multiple orders of magnitude. In Table 8, MMTL reaches the lowest RMSLE at 0.3049 with 19.36%, and MAPLE at 0.0425 with 14.14% improvement compared to the baseline model NN. This indicates that the informative feature embedding (i.e., the feature vector of the last layer of GCN₂) learned from process parameters and defect features in the pre-trained HGCN with abundant defect data in the source task is successfully transferred to critical defects. It significantly improves fatigue life prediction performance with limited data on the target task.

Table 8: Comparison of prediction performance for MMTL against benchmark on the augmented data with all features. Standard deviations for 5-fold cross-validation are presented in parentheses.

		Our model	Benchmark (Pre-trained model + NN layer)			Baseline
		MMTL	HNN+NN layer	GCN+NN layer	NN+NN layer	NN
RMSLE	Test	0.3049 (0.0796)	0.5721 (0.2289)	0.5493 (0.0446)	0.3733 (0.0479)	0.3781 (0.0604)
	Improved %	↑ 19.36	↓	↓	↑ 1.27	-
MAPLE	Test	0.0425 (0.0084)	0.0837 (0.0341)	0.0782 (0.0068)	0.0500 (0.0058)	0.0495 (0.0049)
	Improved %	↑ 14.14	↓	↓	↓	-

Applying MMTL for fatigue life prediction of new L-PBF specimens Using the pre-trained HGCN from thousands of defects for defect classification in the source task of MMTL, we randomly select 80% fatigue data to train the NN for fatigue life prediction. Afterward, we test the MMTL with left 20% data to evaluate the prediction accuracy of MMTL. It is noted that all the predicted fatigue life values of the test set are within the 2 error band (i.e., 2 times the true value, a standard metric to validate the fatigue life prediction in the literature (Shi et al. (2023); Wang et al. (2023))), as shown in Figure 10 (a). Importantly, MMTL can accurately infer the types of critical defects and predict the fatigue life of L-PBF specimens fabricated under different process conditions with given stress amplitudes with a low RMSLE of 0.0270 and MAPLE of 0.0119 (Figure 10 (b-d)).

Discussion on the complexity of the MMTL framework MMTL comprises a total of 2,532 parameters, including 129 parameters from the pre-trained HGCN model and 2,403 parameters from the NN layer, as shown in Table 9. The pre-trained HGCN model consists of a single GCN layer. The simple structure can avoid over-smoothing (i.e., after multiple layers of message passing, node representations become indistinguishable from each other,

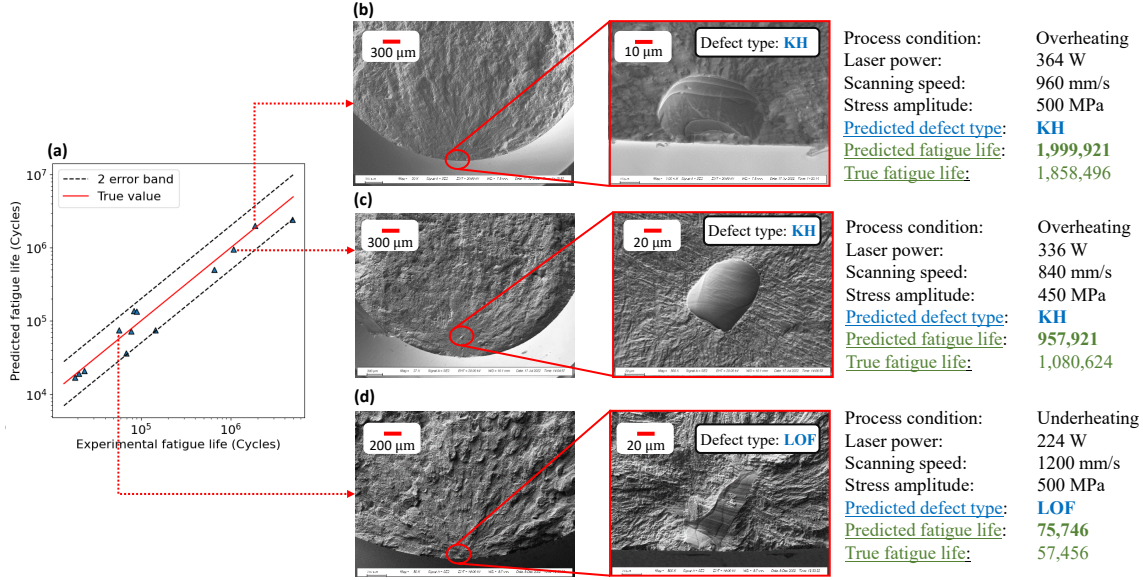


Figure 10: (a) The study compares fatigue life predictions between experimentally obtained results and predictions from MMTL. (b-d) represent the critical defect observed in fractography and MMTL prediction results of three distinct L-PBF fatigue testing specimens. MMTL can accurately predict the fatigue life of specimens in (b), (c), and (d).

losing the ability to capture meaningful differences between nodes) but still can capture the complex local structure of the graph based on node aggregation. The NN layer for fatigue life prediction effectively balances the bias-variance tradeoff in our case with limited data. This layer was selected through cross-validation, given the complex and non-linear relationships between the input features (learned features, loading conditions, and defect distance to the surface) and fatigue life. It demonstrates robust performance across various scenarios in both simulations and the case study, avoiding overfitting and underfitting.

MMTL requires 31.1 seconds to train the HGCN model on a massive dataset and only 4.4 seconds to train the NN layers for fatigue life prediction, compared to 5.3 seconds for the baseline model NN. Although the MMTL demands more time for training, this is performed offline, making the training duration inconsequential. The crucial advantage is that testing new specimens is extremely fast, occurring in milliseconds, similar to the baseline model NN, which is highly suitable for nondestructive fatigue life prediction of L-PBF parts.

Table 9: The number of parameters and training time of compared models in the case study. All models undergo cross-validation to ensure convergence and mitigate both overfitting and underfitting.

		Our model	Benchmark (Pre-trained model + NN layers)			Baseline model
		MMTL	HNN+NN layers	GCN+NN layers	NN+NN layer	NN
Pre-trained model	Parameters	129	89	66	54	-
	Training time (s)	31.1	1.6	5.8	2.1	-
NN layers	Parameters	2403	2403	2403	2403	2851
	Training time (s)	4.4	5.4	7.6	7.1	5.3
Total parameters		2532	2492	2469	2457	2851
Total training time (s)		35.5	7.0	14.4	9.2	5.3

7 Conclusions and Future Work

In this study, we proposed an MMTL framework to enable physics-informed L-PBF fatigue life prediction by incorporating various influencing modalities, such as L-PBF process parameters, XCT-inspected defects, and fatigue loading conditions. Employing HGCN, we represented process and defect data with graphs, leveraged the similarity among defects, and achieved accurate defect classification (F1-score = 0.9593). The learned informative embedding from the process-defect relationship in HGCN was transferred into fatigue life modeling to enable nondestructive prediction of fatigue life for individual L-PBF parts, even with limited data, achieving a RMSLE at 0.3049 with 19.36%, and MAPLE at 0.0425 with 14.14% improvement in fatigue life prediction compared to the baseline model NN. Furthermore, MMTL provides valuable insight of process-defect-fatigue relationships, which can benefit L-PBF process optimization and further adoption of L-PBF parts in engineering applications. Additionally, the demonstrated model efficiency and flexibility indicated the promising potential to extend the MMTL to other areas for prediction with multimodal input data.

For future work, our objectives are to enhance MMTL and broaden its scope of application: (1) enhancing the generalization capability of MMTL. This includes incorporating the shapes and environmental factors of L-PBF parts into modeling to advance nondestructive fatigue life prediction. Such enhancements will account for the increased complexity of loading conditions and loading history in L-PBF parts. (2) exploring more features concerning the intricate relationship between process parameters, defects, crack initiation, and fatigue fracture in L-PBF parts. Our approach will blend empirical data-driven insights with fundamental principles of physics. This entails thoroughly examining the root causes behind

defect formation and how defects and cracks influence fatigue behavior. (3) improving the applicability of MMTL across diverse research domains characterized by limited multimodal data. This investigation will assess the adaptability and effectiveness of MMTL in scenarios where data availability is constrained.

Data Availability Statement

Raw data were generated at the authors' institution. Derived data supporting the findings of this study are available from the corresponding author on request.

References

- Bao, H., Wu, S., Wu, Z., Kang, G., Peng, X., and Withers, P. J. (2021). A machine-learning fatigue life prediction approach of additively manufactured metals. *Engineering Fracture Mechanics*, 242:107508.
- Caton, M., Jones, J. W., Boileau, J., and Allison, J. (1999). The effect of solidification rate on the growth of small fatigue cracks in a cast 319-type aluminum alloy. *Metallurgical and Materials Transactions A*, 30:3055–3068.
- Choo, H., Sham, K.-L., Bohling, J., Ngo, A., Xiao, X., Ren, Y., Depond, P. J., Matthews, M. J., and Garlea, E. (2019). Effect of laser power on defect, texture, and microstructure of a laser powder bed fusion processed 316l stainless steel. *Materials & Design*, 164:107534.
- Criales, L. E., Arisoy, Y. M., Lane, B., Moylan, S., Donmez, A., and Özel, T. (2017). Laser powder bed fusion of nickel alloy 625: Experimental investigations of effects of process parameters on melt pool size and shape with spatter analysis. *International Journal of Machine Tools and Manufacture*, 121:22–36.
- Dong, S., Xiao, J., Hu, X., Fang, N., Liu, L., and Yao, J. (2023). Deep transfer learning based on bi-lstm and attention for remaining useful life prediction of rolling bearing. *Reliability Engineering & System Safety*, 230:108914.
- El Haddad, M., Smith, K., and Topper, T. (1979). Fatigue crack propagation of short cracks.
- Gong, H., Rafi, K., Gu, H., Starr, T., and Stucker, B. (2014). Analysis of defect generation in ti-6al-4v parts made using powder bed fusion additive manufacturing processes. *Additive Manufacturing*, 1:87–98.
- Gordon, J. V., Narra, S. P., Cunningham, R. W., Liu, H., Chen, H., Suter, R. M., Beuth, J. L., and Rollett, A. D. (2020). Defect structure process maps for laser powder bed fusion additive manufacturing. *Additive Manufacturing*, 36:101552.

- Günther, J., Krewerth, D., Lippmann, T., Leuders, S., Tröster, T., Weidner, A., Biermann, H., and Niendorf, T. (2017). Fatigue life of additively manufactured ti-6al-4v in the very high cycle fatigue regime. *International Journal of Fatigue*, 94:236–245.
- Hang, C.-W. and Singh, M. P. (2010). Trust-based recommendation based on graph similarity. In *Proceedings of the 13th international workshop on trust in agent societies (trust). toronto, canada*, volume 82.
- Haridas, R. S., Thapliyal, S., Agrawal, P., and Mishra, R. S. (2020). Defect-based probabilistic fatigue life estimation model for an additively manufactured aluminum alloy. *Materials Science and Engineering: A*, 798:140082.
- Horñas, J., Béhal, J., Homola, P., Doubrava, R., Holzleitner, M., and Senck, S. (2023). A machine learning based approach with an augmented dataset for fatigue life prediction of additively manufactured ti-6al-4v samples. *Engineering Fracture Mechanics*, page 109709.
- Hu, Y., Wu, S., Withers, P., Zhang, J., Bao, H., Fu, Y., and Kang, G. (2020). The effect of manufacturing defects on the fatigue life of selective laser melted ti-6al-4v structures. *Materials & Design*, 192:108708.
- Javidrad, H., Koc, B., Bayraktar, H., Simsek, U., and Gunaydin, K. (2024). Fatigue performance of metal additive manufacturing: a comprehensive overview. *Virtual and Physical Prototyping*, 19(1):e2302556.
- Jia, Y., Fu, R., Ling, C., Shen, Z., Zheng, L., Zhong, Z., and Hong, Y. (2023). Fatigue life prediction based on a deep learning method for ti-6al-4v fabricated by laser powder bed fusion up to very-high-cycle fatigue regime. *International Journal of Fatigue*, 172:107645.
- Kasperovich, G., Haubrich, J., Gussone, J., and Requena, G. (2016). Correlation between porosity and processing parameters in tial6v4 produced by selective laser melting. *Materials & Design*, 105:160–170.
- Kim, F. H., Pintar, A. L., Moylan, S. P., and Garboczi, E. J. (2019). The influence of x-ray computed tomography acquisition parameters on image quality and probability of detection of additive manufacturing defects. *Journal of manufacturing science and engineering*, 141(11):111002.
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kipf, T. N. and Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.
- Komijani, M., Gracie, R., and Sarvaramini, E. (2019). Simulation of induced acoustic emission in fractured porous media. *Engineering Fracture Mechanics*, 210:113–131.
- Li, A., Baig, S., Liu, J., Shao, S., and Shamsaei, N. (2022). Defect criticality analysis on fatigue life of l-pbf 17-4 ph stainless steel via machine learning. *International Journal of Fatigue*, 163:107018.
- Li, Y., Wei, P., Xiang, G., Jia, C., and Liu, H. (2023). Gear contact fatigue life prediction based on transfer learning. *International Journal of Fatigue*, 173:107686.

- Maierhofer, J., Pippan, R., and Gänser, H.-P. (2014). Modified nasgro equation for physically short cracks. *International Journal of fatigue*, 59:200–207.
- Masuo, H., Tanaka, Y., Morokoshi, S., Yagura, H., Uchida, T., Yamamoto, Y., and Murakami, Y. (2017). Effects of defects, surface roughness and hip on fatigue strength of ti-6al-4v manufactured by additive manufacturing. *Procedia Structural Integrity*, 7:19–26.
- McDowell, D., Gall, K., Horstemeyer, M., and Fan, J. (2003). Microstructure-based fatigue modeling of cast a356-t6 alloy. *Engineering Fracture Mechanics*, 70(1):49–80.
- Meneghetti, G., Rigon, D., and Gennari, C. (2019). An analysis of defects influence on axial fatigue strength of maraging steel specimens produced by additive manufacturing. *International journal of fatigue*, 118:54–64.
- Murakami, Y., T., E., K.J., E. M., and de los Rios E.R. (1986). The behavior of short cracks. *EGF publication 1*.
- Nakatani, M., Masuo, H., Tanaka, Y., and Murakami, Y. (2019). Effect of surface roughness on fatigue strength of ti-6al-4v alloy manufactured by additive manufacturing. *Procedia Structural Integrity*, 19:294–301.
- Niu, X., Zhu, S.-P., He, J.-C., Liao, D., Correia, J. A., Berto, F., and Wang, Q. (2022). Defect tolerant fatigue assessment of am materials: Size effect and probabilistic prospects. *International Journal of Fatigue*, 160:106884.
- on Mechanical Testing, A. I. C. E.-. (2016). *Standard test methods for tension testing of metallic materials*. ASTM international.
- Pal, S., Lojen, G., Hudak, R., Rajtukova, V., Brajlilh, T., Kokol, V., and Drstvenšek, I. (2020). As-fabricated surface morphologies of ti-6al-4v samples fabricated by different laser processing parameters in selective laser melting. *Additive Manufacturing*, 33:101147.
- Pegues, J., Shao, S., Shamsaei, N., Sanaei, N., Fatemi, A., Warner, D., Li, P., and Phan, N. (2020). Fatigue of additive manufactured ti-6al-4v, part i: The effects of powder feedstock, manufacturing, and post-process conditions on the resulting microstructure and defects. *International Journal of Fatigue*, 132:105358.
- Peng, X., Wu, S., Qian, W., Bao, J., Hu, Y., Zhan, Z., Guo, G., and Withers, P. J. (2022). The potency of defects on fatigue of additively manufactured metals. *International Journal of Mechanical Sciences*, 221:107185.
- Poudel, A., Yasin, M. S., Ye, J., Liu, J., Vinel, A., Shao, S., and Shamsaei, N. (2022). Feature-based volumetric defect classification in metal additive manufacturing. *Nature Communications*, 13(1):6369.
- Qin, Z., Kang, N., Zhang, F., Wang, Z., Wang, Q., Chen, J., Lin, X., and Huang, W. (2021). Role of defects on the high cycle fatigue behavior of selective laser melted al–mg–sc–zr alloy. *International Journal of Fracture*, pages 1–15.
- Salvati, E., Tognan, A., Laurenti, L., Pelegatti, M., and De Bona, F. (2022). A defect-based physics-informed machine learning framework for fatigue finite life prediction in additive manufacturing. *Materials & Design*, 222:111089.

- Sandoval, C., Malcher, L., Canut, F., Araújo, L., Doca, T., and Araújo, J. (2020). Micromechanical gurson-based continuum damage under the context of fretting fatigue: Influence of the plastic strain field. *International Journal of Plasticity*, 125:235–264.
- Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). Nih image to imagej: 25 years of image analysis. *Nature methods*, 9(7):671–675.
- Senanayaka, A., Tian, W., Falls, T., and Bian, L. (2023). Understanding the effects of process conditions on thermal-defect relationship: A transfer machine learning approach. *Journal of Manufacturing Science and Engineering*, 145(7):071010.
- Shao, S., Mahtabi, M. J., Shamsaei, N., and Thompson, S. M. (2017). Solubility of argon in laser additive manufactured α -titanium under hot isostatic pressing condition. *Computational Materials Science*, 131:209–219.
- Shi, T., Sun, J., Li, J., Qian, G., and Hong, Y. (2023). Machine learning based very-high-cycle fatigue life prediction of als10mg alloy fabricated by selective laser melting. *International Journal of Fatigue*, 171:107585.
- Snell, R., Tammas-Williams, S., Chechik, L., Lyle, A., Hernández-Nava, E., Boig, C., Panoutsos, G., and Todd, I. (2020). Methods for rapid pore classification in metal additive manufacturing. *Jom*, 72:101–109.
- Song, T.-A., Chowdhury, S. R., Yang, F., Jacobs, H., El Fakhri, G., Li, Q., Johnson, K., and Dutta, J. (2019). Graph convolutional neural networks for alzheimer’s disease classification. In *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*, pages 414–417. IEEE.
- Thrun, S. and Pratt, L. (1998). Learning to learn: Introduction and overview. In *Learning to learn*, pages 3–17. Springer.
- Tridello, A., Boursier Niutta, C., Berto, F., Qian, G., and Paolino, D. S. (2021). Fatigue failures from defects in additive manufactured components: A statistical methodology for the analysis of the experimental results. *Fatigue & Fracture of Engineering Materials & Structures*, 44(7):1944–1960.
- Tucho, W. M., Lysne, V. H., Austbø, H., Sjolyst-Kverneland, A., and Hansen, V. (2018). Investigation of effects of process parameters on microstructure and hardness of slm manufactured ss316l. *Journal of Alloys and Compounds*, 740:910–925.
- Wang, L., Zhu, S.-P., Luo, C., Niu, X., and He, J.-C. (2023). Defect driven physics-informed neural network framework for fatigue life prediction of additively manufactured materials. *Philosophical Transactions of the Royal Society A*, 381(2260):20220386.
- Wei, X., Zhang, C., Han, S., Jia, Z., Wang, C., and Xu, W. (2022). High cycle fatigue sn curve prediction of steels based on transfer learning guided long short term memory network. *International Journal of Fatigue*, 163:107050.
- Xiao, L., Xue, X., Wang, N., Ren, Q., Hua, J., and Wang, F. (2023). A transfer learning-based approach to fatigue life prediction of corroded bimetallic steel bars using small samples. *Construction and Building Materials*, 400:132679.

- Xu, R., Huang, S., Song, Z., Gao, Y., and Wu, J. (2023). A deep mixed-effects modeling approach for real-time monitoring of metal additive manufacturing process. *IISE Transactions*, pages 1–15.
- Xue, Y., Horstemeyer, M., McDowell, D., El Kadiri, H., and Fan, J. (2007). Microstructure-based multistage fatigue modeling of a cast ae44 magnesium alloy. *International journal of fatigue*, 29(4):666–676.
- Yamashita, Y., Murakami, T., Mihara, R., Okada, M., and Murakami, Y. (2017). Defect analysis and fatigue design basis for ni-based superalloy 718 manufactured by additive manufacturing. *Procedia structural integrity*, 7:11–18.
- Yao, L., Mao, C., and Luo, Y. (2019). Graph convolutional networks for text classification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 7370–7377.
- Ye, J., Poudel, A., Liu, J., Vinel, A., Silva, D., Shao, S., and Shamsaei, N. (2023). Machine learning augmented x-ray computed tomography features for volumetric defect classification in laser beam powder bed fusion. *The International Journal of Advanced Manufacturing Technology*, 126(7-8):3093–3107.

8 Appendix

8.1 The derivation of HGNC weight updates

We present the derivative of back-propagation to update HGNC model parameters (θ_1 of GCN₁ and θ_2 of GCN₂) as mentioned in Section 4.1 via adaptive moment estimation (Adam). For convenience, we assume these two GCNs share common hyper-parameters (e.g., stepsize α , exponential decay rates γ_1 , and γ_2 for the moments' estimates). Normally, good default settings are $\alpha = 0.001$, $\gamma_1 = 0.9$, $\gamma_2 = 0.999$ and $\epsilon = 10^{-8}$ (Kingma and Ba (2014)). The parameters are updated based on,

$$\begin{aligned}\theta_1^t &= \theta_1^{t-1} - \alpha \cdot \frac{\hat{m}_1^t}{\sqrt{\hat{v}_1^t + \epsilon}} \\ \theta_2^t &= \theta_2^{t-1} - \alpha \cdot \frac{\hat{m}_2^t}{\sqrt{\hat{v}_2^t + \epsilon}}\end{aligned}\tag{9}$$

where $\hat{m}_1^t$ and $\hat{m}_2^t$ are the bias-corrected estimated 1st (mean) and 2nd (uncentered variance) moment vector at timestep t , respectively, to estimate the moment of gradients,

$$\begin{aligned}\hat{m}_1^t &= \frac{\mathbf{m}_1^t}{1 - \gamma_1}, \quad \hat{m}_2^t = \frac{\mathbf{m}_2^t}{1 - \gamma_1} \\ \hat{v}_1^t &= \frac{\mathbf{v}_1^t}{1 - \gamma_2}, \quad \hat{v}_2^t = \frac{\mathbf{v}_2^t}{1 - \gamma_2}\end{aligned}\tag{10}$$

where $\mathbf{m}^t$ and $\mathbf{v}^t$ are the biased 1st and 2nd moment vectors,

$$\begin{aligned}\mathbf{m}_1^t &= \gamma_1 \cdot \mathbf{m}_1^{t-1} + (1 - \gamma_1) \cdot \nabla_{\theta_1} \mathcal{L}^{S,t}(\theta_1^{t-1}, \theta_2^{t-1}) \\ \mathbf{m}_2^t &= \gamma_1 \cdot \mathbf{m}_2^{t-1} + (1 - \gamma_1) \cdot \nabla_{\theta_2} \mathcal{L}^{S,t}(\theta_1^{t-1}, \theta_2^{t-1}) \\ \mathbf{v}_1^t &= \gamma_2 \cdot \mathbf{v}_1^{t-1} + (1 - \gamma_2) \cdot (\nabla_{\theta_1} \mathcal{L}^{S,t}(\theta_1^{t-1}, \theta_2^{t-1}) \odot \nabla_{\theta_1} \mathcal{L}^{S,t}(\theta_1^{t-1}, \theta_2^{t-1})) \\ \mathbf{v}_2^t &= \gamma_2 \cdot \mathbf{v}_2^{t-1} + (1 - \gamma_2) \cdot (\nabla_{\theta_2} \mathcal{L}^{S,t}(\theta_1^{t-1}, \theta_2^{t-1}) \odot \nabla_{\theta_2} \mathcal{L}^{S,t}(\theta_1^{t-1}, \theta_2^{t-1}))\end{aligned}\tag{11}$$

To simplify the derivative of the update for θ_1 and θ_2 , we consider single-layer GCN₁ and GCN₂ since it is simply to expand the update of parameters for multi-layer GCNs based on the derivative chain rule. It means that the number of layers $J = 1$ and $K = 1$ in GCN₁ and GCN₂ as mentioned in Section 4.1. We can rewrite Eq.(4) and Eq.(5) as $g_1(\theta_1; \mathbf{X}_1^S) =$

$\text{ReLU}(\tilde{\mathbf{A}}_1^* \mathbf{X}_1^S \boldsymbol{\theta}_1)$ and $g^{(2)}(g_2(\boldsymbol{\theta}_2; \mathbf{X}_2^S) = \ddot{s} \circ \mathbf{H}_2 = \ddot{s} \circ \text{ReLU}(\tilde{\mathbf{A}}_2^* [\mathbf{X}_2^S | g_1(\boldsymbol{\theta}_1; \mathbf{X}_1^S)] \boldsymbol{\theta}_2)$. Here, we use $\ddot{s}$ to replace Softmax. Now, we solve the $\nabla_{\boldsymbol{\theta}_1} \mathcal{L}^S(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and $\nabla_{\boldsymbol{\theta}_2} \mathcal{L}^S(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ ($\mathcal{L}^S$ is formulated in Eq.(2) and Eq.(6) and we ignore the subscript $t - 1$),

$$\begin{aligned}
\nabla_{\boldsymbol{\theta}_1} \mathcal{L}^S(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &= \frac{\partial \mathcal{L}^S}{\partial \boldsymbol{\theta}_1} \\
&= \frac{\partial \mathcal{L}^S}{\partial g_2(\boldsymbol{\theta}_2)} \cdot \frac{\partial g_2(\boldsymbol{\theta}_2)}{\partial g_1(\boldsymbol{\theta}_1)} \cdot \frac{\partial g_1(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1} \\
&= -\frac{O^S}{g_2(\boldsymbol{\theta}_2)} \cdot \nabla_{\mathbf{H}_2} \ddot{s} \cdot (\boldsymbol{\theta}_2^T \tilde{\mathbf{A}}_2^*) \cdot [\mathbf{0} | \mathbf{I}] \cdot (\tilde{\mathbf{A}}_1^* \mathbf{X}_1^S) \\
&= -\frac{O^S}{\ddot{s}(\mathbf{H}_2)} \cdot \nabla_{\mathbf{H}_2} \ddot{s} \cdot (\boldsymbol{\theta}_2^T \tilde{\mathbf{A}}_2^*) \cdot [\mathbf{0} | \mathbf{I}] \cdot (\tilde{\mathbf{A}}_1^* \mathbf{X}_1^S) \\
&= -O^S \cdot (\mathbf{I} - \ddot{s}(\mathbf{H}_2)) \cdot (\boldsymbol{\theta}_2^T \tilde{\mathbf{A}}_2^*) \cdot [\mathbf{0} | \mathbf{I}] \cdot (\tilde{\mathbf{A}}_1^* \mathbf{X}_1^S) \\
&= -O^S \cdot (\mathbf{I} - \ddot{s}(\text{ReLU}(\tilde{\mathbf{A}}_2^* [\mathbf{X}_2^S | g_1(\boldsymbol{\theta}_1; \mathbf{X}_1^S)] \boldsymbol{\theta}_2)) \\
&\quad \cdot (\boldsymbol{\theta}_2^T \tilde{\mathbf{A}}_2^*) \cdot [\mathbf{0} | \mathbf{I}] \cdot (\tilde{\mathbf{A}}_1^* \mathbf{X}_1^S) \\
&= -\underbrace{(\mathbf{I} - \ddot{s}(\text{ReLU}(\tilde{\mathbf{A}}_2^* [\mathbf{X}_2^S | \text{ReLU}(\tilde{\mathbf{A}}_1^* \mathbf{X}_1^S \boldsymbol{\theta}_1)] \boldsymbol{\theta}_2)))}_{\text{Internal constant part}} \underbrace{\boldsymbol{\theta}_1}_{\text{Parameter part}} \\
&\quad \cdot (\boldsymbol{\theta}_2^T \otimes \tilde{\mathbf{A}}_2^*) O^S \cdot \underbrace{[\mathbf{0} | \mathbf{I}] \cdot (\tilde{\mathbf{A}}_1^* \mathbf{X}_1^S)}_{\text{External constant part}}
\end{aligned} \tag{12}$$

$$\begin{aligned}
\nabla_{\boldsymbol{\theta}_2} \mathcal{L}^S(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &= \frac{\partial \mathcal{L}^S}{\partial \boldsymbol{\theta}_2} \\
&= \frac{\partial \mathcal{L}^S}{\partial g_2(\boldsymbol{\theta}_2)} \cdot \frac{\partial g_2(\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2} \\
&= -\frac{O^S}{g_2}(\boldsymbol{\theta}_2) \cdot \nabla_{\mathbf{H}_2} \ddot{s} \cdot (\tilde{\mathbf{A}}_2^* [\mathbf{X}_2^S | g_1(\boldsymbol{\theta}_1; \mathbf{X}_1^S)]) \\
&= -\frac{O^S}{\ddot{s}(\mathbf{H}_1)} \cdot \nabla_{\mathbf{H}_2} \ddot{s} \cdot (\tilde{\mathbf{A}}_2^* [\mathbf{X}_2^S | g_1(\boldsymbol{\theta}_1; \mathbf{X}_1^S)]) \\
&= -O^S (\mathbf{I} - \ddot{s}(\mathbf{H}_2)) \cdot (\tilde{\mathbf{A}}_2^* [\mathbf{X}_2^S | g_1(\boldsymbol{\theta}_1; \mathbf{X}_1^S)]) \\
&= -O^S (\mathbf{I} - \ddot{s}(\text{ReLU}(\tilde{\mathbf{A}}_2^* [\mathbf{X}_2^S | g_1(\boldsymbol{\theta}_1; \mathbf{X}_1^S)] \boldsymbol{\theta}_2))) \\
&\quad \cdot (\tilde{\mathbf{A}}_2^* [\mathbf{X}_2^S | g_1(\boldsymbol{\theta}_1; \mathbf{X}_1^S)]) \\
&= -O^S (\mathbf{I} - \ddot{s}(\text{ReLU}(\tilde{\mathbf{A}}_2^* [\mathbf{X}_2^S | \text{ReLU}(\tilde{\mathbf{A}}_1^* \mathbf{X}_1^S \boldsymbol{\theta}_1)] \boldsymbol{\theta}_2))) \\
&\quad \cdot (\tilde{\mathbf{A}}_2^* [\mathbf{X}_2^S | \text{ReLU}(\tilde{\mathbf{A}}_1^* \mathbf{X}_1^S \boldsymbol{\theta}_1)])
\end{aligned} \tag{13}$$

8.2 Simulation and case study model setup

8.2.1 Simulation model setup

Pre-trained HGCN: we build a hierarchical structure comprising two hierarchies. Hierarchy 1 is responsible for embedding and includes an input layer with neurons matching the number of features in modality 1, a GCN layer with 3 neurons, a hidden layer with 5 neurons and an output layer with 1 neuron. Hierarchy 2 focuses on classification, starting with an input layer that has neurons equal to the sum of features from modality 1 and the embedding features (1, as specified in our paper). It includes a hidden layer with 5 neurons, a GCN layer with 5 neurons, and an output layer with neurons corresponding to the number of classes. The recommended number of training epochs is 500 or 1000.

Pre-trained HNN: we design the HNN, which has a hierarchical structure with two hierarchies. Hierarchy 1 is responsible for embedding and includes an input layer with neurons matching the number of features in modality 1, a hidden layer with 5 neurons, and an output layer with 1 neuron. Hierarchy 2 focuses on classification, starting with an input layer that has neurons equal to the sum of features from modality 1 and the embedding features (1, as specified in our paper). It includes a hidden layer with 5 neurons and an output layer with neurons corresponding to the number of classes. The recommended number of training epochs is 1000.

Pre-trained GCN: we design the GCN, which has an input layer with neurons matching the number of features, a hidden layer with 3 neurons, a GCN layer with 3 neurons, and an output layer with neurons corresponding to the number of classes. The recommendation of training epochs is 1000.

Pre-trained NN: we design the NN, which has an input layer with neurons matching the number of features, a hidden layer with 5 neurons, and an output layer with neurons corresponding to the number of classes. The recommendation of training epochs is 1000.

Retrained NN layers: we design a retrained NN with an input layer with neurons matching the number of features in modality 3 and the embedding features from the pre-trained

model, a hidden layer with 64 neurons, a hidden layer with 32 neurons, and an output layer with 1 neuron. The recommendation of training epochs is 15000.

Baseline NN: we design a baseline model NN without using TL with an input layer with neurons matching the number of features, a hidden layer with 64 neurons, a hidden layer with 32 neurons, and an output layer with 1 neuron. The recommendation of training epochs is 5000.

8.2.2 Case study model setup

Pre-trained HGCN: we build a hierarchical structure comprising two hierarchies. Hierarchy 1 is responsible for embedding and includes an input layer with 2 neurons, a GCN layer with 5 neurons, a hidden layer with 5 neurons and an output layer with 1 neuron. Hierarchy 2 focuses on classification, starting with an input layer that has neurons equal to 8 and the embedding features (1, as specified in our paper). It includes a hidden layer with 5 neurons, a GCN layer with 5 neurons, and an output layer with 3 neurons. The recommended number of training epochs is 1000.

Pre-trained HNN: we design the HNN, which has a hierarchical structure with two hierarchies. Hierarchy 1 is responsible for embedding and includes an input layer with 2 neurons, a hidden layer with 5 neurons, and an output layer with 1 neuron. Hierarchy 2 focuses on classification, starting with an input layer that has neurons equal to 8 and the embedding features (1, as specified in our paper). It includes a hidden layer with 5 neurons and an output layer with 3 neurons. The recommended number of training epochs is 1000.

Pre-trained GCN: we design the GCN, which has an input layer with 10 neurons, a hidden layer with 3 neurons, a GCN layer with 3 neurons, and an output layer with 3 neurons. The recommendation of training epochs is 1000.

Pre-trained NN: we design the NN, which has an input layer with 10 neurons, a hidden layer with 4 neurons, and an output layer with 3 neurons. The recommendation of training epochs is 1000.

Retrained NN layers: we design a retrained NN with an input layer with neurons matching the number of features in modality 3 and the embedding features from the pre-trained model, a hidden layer with 64 neurons, a hidden layer with 32 neurons, and an output layer with 1 neuron. The recommendation of training epochs is 3500.

Baseline NN: we design a baseline model NN without using TL with an input layer with neurons matching the number of features, a hidden layer with 64 neurons, a hidden layer with 32 neurons, and an output layer with 1 neuron. The recommendation of training epochs is 6000.