Physics-informed Hierarchical Learning for Defect Characterization and Fatigue Life Prediction in Metal Additive Manufacturing

by

Anyi Li

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 8, 2026

Keywords: physics-informed hierarchical modeling, machine learning, transfer learning, federated learning, defect characterization, fatigue life prediction, laser powder bed fusion, additive manufacturing

Approved by

Dr. Aleksandr Vinel, Chair, Associate Professor in Industrial and Systems Engineering
Dr. Jia “Peter” Liu, Co-chair, Associate Professor in Industrial and Systems Engineering
Dr. Gregory Harris, Distinguished Professor in Industrial and Systems Engineering
Dr. Shuai Shao, WcWane Professor in Mechanical Engineering

Abstract

Understanding and accurately predicting the fatigue life of additive manufactured (AM) metal parts remains a pressing challenge for their reliable adoption in safety-critical applications such as aerospace, energy, and biomedical systems. In laser powder bed fusion (L-PBF), process-induced volumetric defects are one of the primary factors governing crack initiation and fatigue failure. Traditional fatigue assessment relies on destructive and time-consuming fatigue testing, resulting in fatigue data that is expensive and scarce. Consequently, individual organizations often possess insufficient data to develop robust predictive models. While collaborative learning across organizations provides a potential solution, concerns about heterogeneous inspection capabilities and data privacy hinder effective data sharing and joint model development. Moreover, most existing studies focus on uniaxial loading conditions, whereas engineering components are usually experienced complex multiaxial stress states. These challenges span various levels of fatigue assessment, ranging from defect-level characterization and specimen-level fatigue life prediction to organization-level collaborative learning and multiple-defect level multiaxial fatigue behavior understanding.

To address these challenges, this research develops a physics-informed hierarchical learning framework for defect characterization, fatigue life prediction, and collaborative learning in AM. The framework progressively enhances defect-fatigue understanding and prediction across multiple hierarchical levels through a series of physics-informed data-driven methodologies. Specifically, four methodologies are introduced:

(1) *An integrated data-driven analytical framework using kernel support vector regression and interpretable model-agnostic methods for defect criticality analysis and fatigue life prediction of L-PBF 17-4 PH stainless steel parts.* As the foundation of the hierarchical framework, this defect-level methodology enhances the understanding of defect-fatigue relationships in a data-driven manner and quantitatively analyzes the importance of defect features to fatigue life.

(2) *A multimodal transfer learning (MMTL) framework using process-defect-loading data for defect classification and nondestructive fatigue life prediction of L-PBF Ti-6Al-4V parts.* At the specimen level, this framework leverages knowledge learned from 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.

(3) *A personalized federated transfer learning framework using conditional optimal transport (FedCOT) for collaborative predictive modeling across organizations with heterogeneous inspection capabilities.* At the organization level, this framework enables "target" organizations with limited features to benefit from "source" organizations with sufficient features in terms of prediction performance, while preserving data privacy through a central server.

(4) *A physics-informed multiple instance learning (PhysMIL) framework for multiaxial fatigue life prediction of AM metal parts.* At the multiple-defect level, this framework integrates the critical plane mechanics with defect representations from multiple defects within each specimen and learns their contributions to fatigue failure via attention-based aggregation and physics-informed regularization on defect characteristics and fatigue damage mechanisms. It improves both prediction accuracy and interpretability under complex multiaxial loading conditions.

These methodologies establish a unified physics-informed hierarchical learning framework that advances the understanding of process-defect-fatigue relationships, enables nondestructive and data-efficient fatigue life prediction, supports privacy-preserving collaboration across organizations, and extends predictive modeling to realistic multiaxial loading scenarios. They provide practical tools for improving the reliability and qualification of AM components and can be utilized in other engineering domains involving limited, heterogeneous, and privacy-sensitive data.

Artificial Intelligence (AI) Use Disclosure Statement

In the preparation of this dissertation, the following Artificial Intelligence (AI) tools were used: ChatGPT, Claude, and Google Gemini. These tools were used primarily for language editing, manuscript organization, grammar improvement, code debugging assistance, and brainstorming of research ideas. The author acknowledges full responsibility for the intellectual content of this work and has ensured that all AI-assisted sections have been reviewed and revised for accuracy and appropriate academic style. All AI-generated content was reviewed and validated for relevance, appropriateness, and accuracy before incorporation into the final document to maintain the scholarly integrity of this research.

Digital Accessibility Disclosure Statement

In the preparation of this dissertation, the following digital accessibility tools were used to ensure this document complies with federal requirements: Microsoft Word Accessibility Checker, Grammarly, and university-provided electronic thesis and dissertation (ETD) accessibility guidelines. The author acknowledges full responsibility for the intellectual content of this work and has made a good faith effort to comply with digital accessibility requirements in publishing, wherein the nature of the content does not significantly change in order to do so. Furthermore, all content has been reviewed and revised to meet these requirements prior to final publication.

Acknowledgments

Time flies, and everything changes, yet it is through change that we discover who we are and what truly matters. Looking back on this journey, I realize that the passage of time is measured not only by years but also by the experiences, relationships, and lessons that shape us along the way. I came into this world to experience life, and the Ph.D. journey has been both fast and slow. Fast in the passing of time, yet slow in the accumulation of knowledge, lessons, and wisdom. The challenges, setbacks, and failures have shaped who I am today and have been one of the most valuable lessons of my life.

First, I would like to express my deepest gratitude to my family—my mother and father, and especially my wife, Xiaoling, and our beloved dog, Turbo. Their love, support, and companionship have been a constant source of strength throughout this journey. During the difficult moments, they were the light that pierced through the darkness and abyss and gave me the courage to keep moving forward.

I am truly grateful to my advisor, Dr. Liu. In the past few years, we fought side by side. I have learned from him far more than research methodologies and technical knowledge. He made me realize how to think critically, pursue meaningful research, and develop taste and judgment to become an independent scholar and scientist. More importantly, I learned how to be a responsible person who strives to advance science while remaining devoted to family and life. Words cannot fully express my appreciation for his kind mentorship, guidance, and trust.

I would also like to express my heartfelt thanks to my co-advisor, Dr. Vinel, for his support throughout my Ph.D. journey. I enjoyed our insightful discussion, his patience, encouragement, and meticulous advice, which played an important role in my academic development.

My sincere appreciation also goes to Dr. Shao, Dr. Shamsaei, Dr. Cheng, and Dr. Harris for their invaluable collaboration, guidance, and constructive feedback over the years. Their expertise, suggestions, and high standards have greatly shaped my growth as both a researcher

and collaborator.

I am also grateful to all my colleagues and friends who shared this journey with me. Special thanks to Rupesh, Tianyu, Bowen, Shehzaib, Qais, Amir, Jie, Jiafeng, and many others who have been part of this unforgettable chapter of my life. Best wishes to you!

Finally, I would like to say that the Ph.D. is not merely a degree but a journey of self-discovery, self-doubt, repeatedly rebuilding self-confidence, perseverance, and transformation. I will carry the lessons, memories, and relationships for the rest of my life.

Life is only lived once, so embrace it, cherish it, enjoy it, and live it to the fullest. Be loving to family, passionate to the world, and resilient to yourself.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ALE	Accumulated local effects
AM	Additive manufacturing
CBIR	Content-based image retrieval
COT	Conditional optimal transport
CT	Computed tomography
DTS	The distance to the surface
ED	Equivalent diameter
FedCOT	Personalized Federated Transfer Learning via Conditional Optimal Transport
FL	Federated learning
HGCN	Hierarchical graph convolutional network
IDADC	Integrated data-driven analytical framework for defect criticality
L-PBF	Laser beam powder bed fusion
LASSO	Least absolute shrinkage and selection operator
LOFs	Lack of fusions
MajorAL	Major axis length
MAPE	Mean absolute percentage error
MaxFD	Max Feret diameter

MIC	Maximal information coefficient
MIL	Multiple instance learning
MinFD	Min Feret diameter
MinorAL	Minor axis length
ML	Machine learning
MMTL	Multimodal transfer learning
MSE	Mean square error
NSL	Normalized stress level
PCC	Pearson correlation coefficient
PFI	Permutation Feature Importance
PhysMIL	Physics-informed multiple instance learning
PSWT	Smith–Watson–Topper fatigue parameter
RMSE	Root mean square error
Sa	Strain amplitude
SEM	Scanning electron microscopy
SSE	Sum of square error
SVR	Support vector regression
UTS	Ultimate tensile strength
XCT	X-ray computed tomography
YS	Yield strength

Chapter 1

Introduction

1.1 Background and Motivation

Additive manufacturing (AM) fabricates parts in a layer-wise manner, which has transformed from a rapid prototyping tool into a revolutionary digital manufacturing technology for functional and structural parts [1]. Among many AM processes, laser powder bed fusion (L-PBF) has dominated the metal AM market because of its strong capability to fabricate complex geometries, customized products, and low-volume production [2]. This makes L-PBF a promising process for critical manufacturing applications in aerospace, biomedical, and energy sectors.

However, the uncertainty of fatigue performance in L-PBF parts is a key challenge for the adoption of L-PBF in high-stakes applications, such as aerospace and biomedical fields. The fatigue behavior of L-PBF parts is affected by microstructure, residual stress, surface roughness, and internal defects. Among these factors, process-induced volumetric defects, inevitably generated due to complex local process dynamics, are particularly critical since they can act as stress concentrators and crack initiation sites, and cause the final failure. Other factors can be effectively mitigated by post-processing methods such as heat treatment or surface polishing. Destructive fatigue testing is an effective method for assessing performance, but it is time-consuming and costly, typically taking days or weeks to complete for a single specimen in high-cycle testing. To enhance the defect-tolerant design and reliable application of L-PBF parts, accurately predicting fatigue life is crucial.

Therefore, there is a growing need to develop physics-informed data-driven approaches that can leverage defect information to predict fatigue life accurately and effectively. These approaches should not only achieve accurate fatigue life prediction under limited fatigue data but also provide physical insights linking processes, defects, and fatigue behavior. Furthermore, fatigue assessment in AM involves various interconnected levels, including defect-level characterization and specimen-level fatigue life prediction to organization-level collaborative learning and multiple-defect level multiaxial fatigue behavior understanding. This dissertation develops a physics-informed hierarchical learning framework that advances fatigue understanding and prediction, supports collaborative learning across organizations with limited data and inspection capability, while preserving data privacy, especially for the small and medium-sized manufacturers (SMM) [3].

1.2 Research Gaps

Despite significant progress in fatigue performance assessment of L-PBF parts, current approaches normally model only specific aspects of the fatigue problem, such as process-defect relationships, defect characterization, fatigue life prediction, etc. However, fatigue assessment in AM spans multiple interconnected levels. Consequently, they are still facing grave challenges:

(1) **Insufficient physics understanding of defect criticality in fatigue failure.** Although volumetric defects are recognized as dominant crack initiation sites, their relative importance and interactions with fatigue life remain unclear. Existing studies usually rely on simple defect features, such as defect size or projected area, which may not fully capture the effects of defect morphology and location. As a result, there is a need for developing interpretable data-driven methodologies that can extract meaningful defect characteristics and explore the defect-fatigue relationships.

(2) **Limited fatigue data availability for robust predictive modeling.** Fatigue testing is destructive, time-consuming, and expensive, resulting in only a limited amount of data. Such limited datasets pose a challenge for predictive models to capture the intrinsic variability of L-PBF processes and to develop reliable fatigue life prediction models. It motivates the development of data-efficient learning approaches that can leverage sufficient information from process and defect data.

(3) **Lack of privacy-preserving collaborative modeling across organizations with**

heterogeneous data sources. Individual organizations (e.g., SMM) often have insufficient fatigue or inspection data to develop robust predictive models independently. Collaborative learning is an efficient way to improve model performance by leveraging data from multiple organizations. However, differences in inspection capabilities bring heterogeneous feature spaces, while privacy and proprietary constraints prevent direct data sharing. These challenges hinder effective use of knowledge sharing and the development of scalable collaborative fatigue life prediction frameworks.

(4) **Limited physics-informed predictive modeling of multiaxial fatigue considering multiple defects.** Current machine learning models have shown promising predictive capabilities for fatigue life prediction; nevertheless, many models rely mainly on data-driven knowledge and do not consider established fatigue mechanics principles. Furthermore, most existing fatigue studies for AM parts focus on uniaxial loading conditions and assume that fatigue behavior is governed by a single critical defect. In practice, engineering components frequently experience complex multiaxial loading, and multiple defects might jointly impact crack initiation and fatigue failure. These challenges motivate the development of physics-informed learning approaches that integrate multiple defect information with established multiaxial fatigue mechanisms.

1.3 Research Objectives

Motivated by the aforementioned challenges, this dissertation develops a physics-informed hierarchical learning framework for defect characterization, fatigue life prediction, and collaborative learning in L-PBF parts. The overall goal is to improve the understanding of process-defect-fatigue mechanisms, enhance fatigue life prediction accuracy, and enable robust predictive learning across organizations under limited, heterogeneous, and privacy-sensitive data environments. In pursuit of this goal, the objectives are specialized into four tasks:

(1) **To develop interpretable data-driven approaches for defect characterization, criticality analysis, and fatigue life prediction**, capturing the geometric, morphological, and location effects of volumetric defects on stress concentration and crack initiation, then improving

fatigue life prediction of AM parts (Task 1).

(2) **To build a multimodal transfer learning framework for fatigue life prediction of L-PBF parts**, integrating process parameters, XCT-inspected defect features, and fatigue loading conditions into a unified predictive framework. This task aims to predict fatigue life nondestructively with limited fatigue test data by leveraging knowledge from abundant process and XCT defect data (Task 2).

(3) **To design a scalable and privacy-preserving federated transfer learning framework**, enabling knowledge sharing among organizations with heterogeneous inspection capabilities while improving overall fatigue life prediction without sharing raw data (Task 3).

(4) **To develop a physics-informed multiple instance learning framework for multiaxial fatigue life prediction**, integrating critical-plane fatigue mechanics with defect representations from multiple defects within each specimen. This task aims to improve fatigue life prediction accuracy and interpretability under realistic multiaxial loading conditions (Task 4).

These methodologies establish a unified physics-informed hierarchical learning framework for modeling process-defect-fatigue relationships and improving fatigue life prediction in AM. The framework progresses from defect criticality analysis and uniaxial fatigue modeling to privacy-preserving collaborative learning and multiaxial fatigue modeling, providing a scalable pathway toward the reliable qualification and deployment of AM components in safety-critical applications. The structure of this dissertation is shown in Figure 1.1.

1.4 Dissertation Organization

The rest of the dissertation is organized as follows: Chapter 2 reviews the related literature on defect characterization and analysis, fatigue behavior modeling, and federated learning for collaboration among organizations. Chapter 3 introduces an interpretable machine learning framework for defect characterization and fatigue life prediction of L-PBF parts. Chapter 4 develops a multimodal transfer learning framework that integrates process parameters, XCT-inspected defects, and fatigue test conditions to enable nondestructive fatigue life prediction with limited

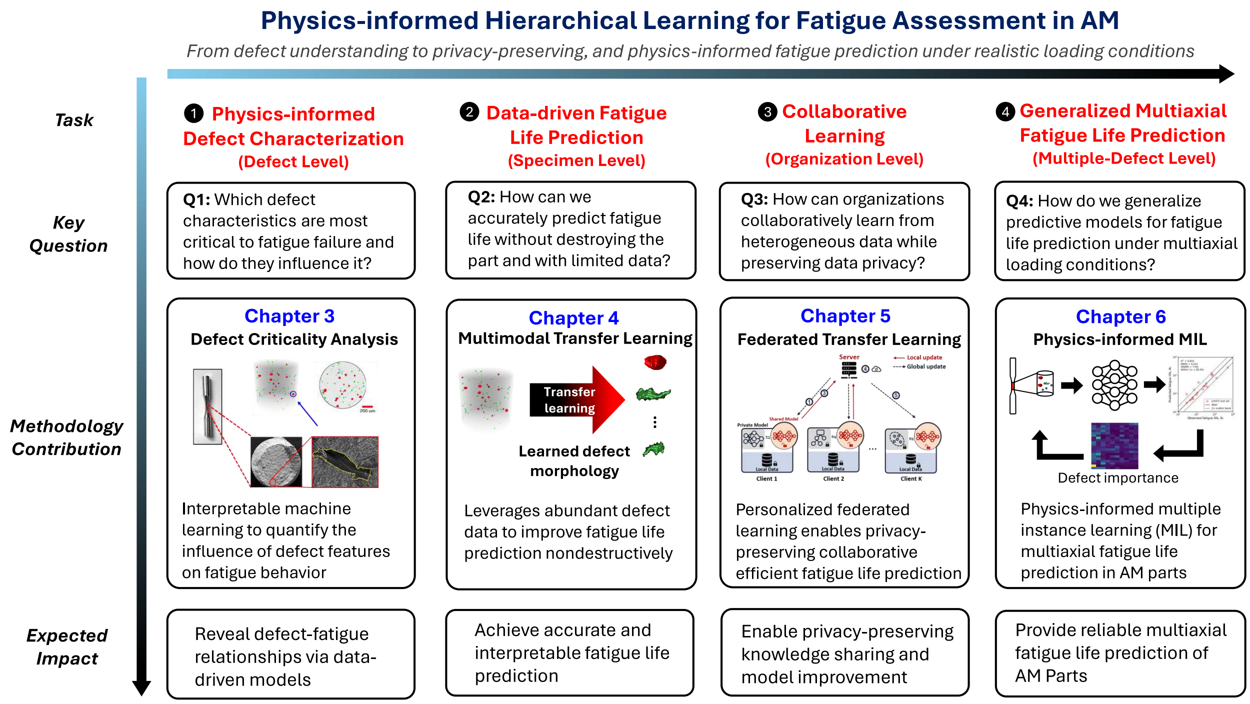

Figure 1.1: Research framework of this dissertation. The work addresses four key questions through a progression of physics-informed defect characterization, data-driven fatigue life prediction, collaborative learning, and generalized multiaxial fatigue life prediction.

data. Chapter 5 presents a scalable and privacy-preserving collaborative framework via federated transfer learning, facilitating knowledge sharing among organizations with heterogeneous data sources without sharing raw information. Chapter 6 proposes a physics-informed multiple instance learning framework for multiaxial fatigue life prediction. The framework integrates the critical plane approach with defect representations from multiple defects to improve prediction accuracy and interpretability under complex loading conditions. Finally, Chapter 7 summarizes the key contributions and findings and discusses future research plans.

Chapter 2

Literature Review

2.1 Defect Characterization and Analysis in Metal Additive Manufacturing

2.1.1 Process-defect relationship in L-PBF

The different types of defect generation in L-PBF Ti-6Al-4V specimens are influenced by the energy input, which is controlled by process parameters such as laser power, scanning speed, hatch distance, and layer thickness [4]. Gong et al. [5] 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. [6] showed that KHs dominated the specimen even with high density in the excessive heat energy window (i.e., over melting). Pal et al. [7] 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 [8]. 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 [9]. In summary, 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 [10]. LOFs, with irregular shapes, are generated by insufficient energy input, leading to significant scatter in fatigue life [11].

2.1.2 Defect inspection, feature extraction, and relationship with fatigue life

Both destructive and nondestructive inspections can be used to characterize defects in AM-fabricated specimens. Destructive defect inspections typically involve fractography analysis, conducted using scanning electron microscopy (SEM) on failed fatigue specimens, to determine the defects responsible for crack initiation and specimen failure. Nondestructive defect inspections, such as X-ray CT techniques, can acquire 3D descriptions of defects within the scanned regions. Nondestructive defect inspections, such as X-ray CT techniques, can acquire 3D descriptions of defects within the scanned regions. Currently, defect types (e.g., LOFs, gas-entrapped pores, or keyholes) and a few conventional defect characteristics (e.g., size/area, circularity/sphericity, aspect ratio, and distance to the surface) from the fractography and X-ray CT have been evaluated to understand their impacts on the fatigue performance of metal AM components [12–16].

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 [10]. 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) [10, 17]. KHs 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 (KHs, LOFs, and GEPs) in L-PBF specimens. Poudel et al. [10] trained decision trees and neural networks to classify defects, and they achieved $> 98\%$ and $> 99\%$ classification accuracy, respectively.

Ye et al. [18] 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. [19] 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. For example, studies using X-ray CT to examine defects on L-PBF fabricated specimens and investigate the effects of defects on mechanical properties were reviewed in [14]. It showed that fatigue properties are more critical to total porosity extent and proximity of defects to the surface, and the use of X-ray CT and the advances in the quality of L-PBF materials have led to the identification of near-surface pores as critical for fatigue applications.

These studies specifically obtain the defect features, such as size, shape, and distance to the surface, from 2D microscopy or 3D X-ray CT images to investigate their relationships with fatigue performance. For instance, Tang and Pistorius [13] examined defects measured on 2D sections of L-PBF AlSi10Mg parts by scanning electron microscope (SEM) and estimated the distribution of defect size by extreme-value statistics. The defect size distribution was used to infer the size of the feasible largest defect and establish the relationships between the size of the largest defect and the fatigue life for prediction. Sanaei et al. [14] studied defects of L-PBF Ti-6Al-4V and 17-4 PH stainless steel (SS) specimens and the relationships between the defect characteristics (e.g., size, sphericity/circularity, aspect ratio) and fatigue life. Based on the defect description with both 2D microscope and 3D X-ray CT, it was noted that the aspect ratio and circularity/sphericity of the defects typically decreased as the defect size increased; the largest defects had low circularity/sphericity and aspect ratio and created high stress-concentrations which made them critical to the fatigue performance of specimens. The variability of defect characteristics based on their location in these specimens was also studied. Tammas-Williams et al. [15] identified critical defects by ranking porosity defects in electron beam-powder bed fusion (EB-PBF) fabricated Ti-6Al-4V specimens according to four defect features: size, aspect ratio, distance to the surface, and distance to other pores from X-ray CT scans. According to authors-defined ranking rules, the defects with high ranks (i.e., large pore size, small aspect ratio, small distance to the surface, and/or small distance to other pores) were considered detrimental to fatigue life. The experiments verified that the actual crack-initiating defects were among the top 3% defects in their rankings. Maskery et al. [16] used X-ray CT to quantify and characterize defects in L-PBF AlSi10Mg parts in terms of size, aspect ratio, and circularity. The data showed that the largest defects in the L-PBF material were strongly anisotropic, flat, and disc-like parallel to the layers of the manufacturing process. Moreover, the defect size, aspect ratio, and circularity were unaffected by the heat treatments, despite the changes in microstructure and hardness.

The abundance and variability of L-PBF process-induced defects can cause degraded fatigue strength and a pronounced scatter in the fatigue life of the fabricated components [20]. Moreover, a systematic investigation and analysis of the defects on fatigue performance remain unexplored, and therefore, defect features (such as morphological features) have not yet been widely used in conventional fracture-mechanics-based fatigue modeling for L-PBF parts.

2.2 Fatigue Behavior Modeling in Metal AM

2.2.1 Conventional and mechanical fatigue models

The fatigue modeling of metallic materials can be classified into three main categories, namely: empirical, analytical, and computational methods.

- Empirical models, such as the Coffin-Manson and Basquin relations, analyze fatigue parameters from experimental data through curve fitting [21, 22]. These methods are simple and accurate but do not explicitly consider the effects of surface roughness and defects inherent in AM materials; new data must be generated whenever defect conditions change.

- Analytical methods, such as fracture-mechanics-based fatigue models, allow fatigue strength to be evaluated with the assumption that cracks or flaws exist, and that fatigue life is governed by crack propagation [23]. Defect-sensitive models have been applied to correlate defect features with AM fatigue performance [24–26]. Commonly adopted hypotheses are that defects behave as "short cracks" and that fatigue limit is inversely related to crack size [27, 28]. Stern et al. [29] used microfocus CT to correlate 3D process-induced defects with fatigue behavior, showing fatigue life is significantly affected by defect size and orientation. The current popular assessment framework includes estimating defect size by extreme-value statistics with a Gumbel distribution, relating it to fatigue limit via Murakami's $\sqrt{area}$ and El-Haddad's model, and determining the infinite life regime with the Kitagawa–Takahashi diagram. In the finite life regime, fatigue life can be calculated by assuming the defect as the initial crack and applying Paris' law or NASGRO [30–33]. Kotzem et al. [34] also adopted the $\sqrt{area}$ model to study the relation between defect size and S–N curves using artificial defects.

- Computational models, such as crystal plasticity simulations, explicitly model deformation mechanisms involving defects and phases [35–39]. Fatigue indicator parameters like Fatemi–Socie, cumulative plastic strain, or strain energy are used to monitor damage evolution [39, 40]. These computations are powerful but limited in length scale and struggle to assess stable long crack growth.

In addition, defect-based semi-empirical models have been widely used. Murakami's $\sqrt{area}$ model relates fatigue life to defect size and hardness [41], applied across alloys including Ti-6Al-4V [42, 43], and extended to improve short-crack growth predictions in NASGRO [44], predict lower fatigue bounds [45], and estimate fatigue life [43, 46]. Defect-tolerant design using the Kitagawa–Takahashi diagram, microstructure-based models, and multistage fatigue (MSF) models have also been used to describe smooth transitions between crack growth regimes and account for incubation and small crack propagation [47–51]. Continuous damage models and stress-life curve approaches are further explored in [52–55].

2.2.2 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. [56] 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. [57] 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. [58] 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 [59]. Limited data can cause the ML model to overfit during training, reducing its ability to accurately predict unseen fatigue life beyond the training set [60].

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 [61, 62]. Researchers have applied TL in various fatigue-related applications. Li et al. [63] 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 an MAPE of 0.4633. Wei et al. [64] 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. [65] 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. [66] 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.

2.2.3 Multiaxial fatigue life prediction in AM

Accurately understanding and predicting multiaxial fatigue life is crucial for evaluating the performance of AM metal parts under complex service loading conditions with normal shear stresses. Traditional multiaxial fatigue models are typically extensions of conventional material criteria and can be categorized into Von Mises and maximum principal stress (MPS) criteria, critical plane approach, and Fracture mechanics approach [67]. The widely used approaches are the critical plane approach (e.g., Fatemi-Socie (FS) parameters [40], Smith-Watson-Topper (SWT) parameters [68]) and the Fracture mechanics approach (e.g., Hartman-Schijve variant of NASGRO equation [69]). The critical plane approach leverages basic material fatigue properties to decide the failure orientation and fatigue life [70, 71]. These approaches are valuable, but do not consider defect geometry, complex stress states, and their interaction.

Recently, statistical and machine learning methods have been explored to address the limitations of conventional multiaxial fatigue models for conventional materials [67, 72–83]. Yang et al. [72] developed a deep learning framework for predicting the multiaxial fatigue life of conventional manufactured parts made from materials such as steel, aluminum, titanium alloy, and natural rubber. Their method incorporates multiaxial stress-strain histories as input features and leverages a Long Short-Term Memory (LSTM) network and a fully connected network to learn a nonlinear relationship between loading paths and fatigue life for each material. Their results showed that their model achieved good prediction accuracy across all materials compared with the traditional FS and SWT models from other literature. Pałczyński et al. [73] evaluated multiple machine learning methods to predict the multiaxial fatigue life of PA38-T6 aluminum alloy, including neural networks, support vector regression (with linear and radial basis functions), decision tree, random forest, and XGBoost methods. They showed that SVR and GPR provided the most accurate and robust predictions, with lower generalization errors compared to empirical models, such as Fatemi-Socie, and Ince-Glinka. Foti et al. [67] worked on a comprehensive literature review of multiaxial fatigue in AM metals, summarizing the traditional fatigue model and indicating the potential to adopt data-driven models to relate several process parameters with the component mechanical behavior to increase their reliability.

Although these studies show the potential of ML for multiaxial fatigue life prediction, most existing approaches focus on a single dominant critical defect, loading histories, stress-strain descriptors, or specimen-level features, and are mainly developed for conventionally manufactured materials. For AM metals, fatigue behavior is influenced by the interaction of multiple defects with different sizes, morphology, and location. Existing ML models typically aggregate defect information via statistical measures, such as the average defect characteristics [84], thereby ignoring the contributions of individual defects. As a result, accurately identifying critical defects and incorporating their contributions into fatigue life prediction is still an open challenge for multiaxial fatigue assessment of AM parts.

2.3 Federated Learning and Optimal Transport

2.3.1 Foundations of FL and data heterogeneity problem

FL is a decentralized machine learning method where multiple clients collaboratively train a global model without sharing their local data [85–88]. Unlike traditional distributed learning, which primarily focuses on parallelizing computations over independent and identically distributed (IID) data, FL can operate under the data heterogeneity that client data is non-IID and varies in size since clients collect and store data independently. This heterogeneity can lead to biased models or slow convergence if not properly addressed.

This work categorizes data heterogeneity in FL into two primary forms. The first is statistical heterogeneity, where clients' collected data follow different distributions due to varying environments or sampling biases [89]. The second is feature heterogeneity [90], where clients measure different sets of features, often caused by differences in sensor deployment, inspection capabilities, data collection protocols, or privacy restrictions.

To address such challenges, researchers usually work on three types of FL:

(1) Horizontal FL (HFL) tackles statistical heterogeneity under a shared feature space, widely studied in large-scale applications [91, 92] (e.g., Google's Gboard for next-word prediction

[93, 94]). Statistical heterogeneity has been extensively studied in the HFL community (e.g., regularization via FedProx and FedEntropy [95, 96], multi-task learning via Ditto and MOCHA [97, 98], public data use via FedMD [99], or clustering via FLACC [100]. More details can be found in the literature [89, 101]. These techniques stabilize training and adapt to heterogeneous data distributions, but they assume all clients have the same feature set. Therefore, they are not feasible when clients observe different features or modalities, and feature heterogeneity remains significantly more challenging and realistic [90].

(2) Vertical FL (VFL) has been developed for scenarios like healthcare and finance, which allows multiple clients to collect different features but share the same user identifiers [91, 102]. Typical methods rely on vertical splitVFL [103] and aggVFL [91, 104] to allow joint training without exposing raw features [102]. However, VFL fundamentally requires aligned sample IDs across clients, which is often impractical in manufacturing. For example, different facilities may fabricate or test entirely different parts even though using the same materials or process settings due to process instability [105].

(3) Federated transfer learning (FTL) is designed for clients with partially overlapping feature and sample spaces [91, 106]. Previous frameworks combined transfer learning with FL to enable knowledge transfer between domains with limited overlap. SFTL [107] trains the encoders to map heterogeneous feature spaces into a common latent subspace through partially aligned samples by minimizing the latent representation distance. Then the target client trains its own model based on the private data. SFHTL [108] adopts an autoencoder to learn local representations and build global representations. The labels of the source client are propagated to each target client. PrADA [109] transfers knowledge from a label-rich source client to the target client via the adversarial domain adaptation that minimizes the difference between the source and target client representations. Yet, these methods depend on overlap data or features (e.g., a subset of common features or a small set of aligned samples) to guide the knowledge transfer [110].

In this case, where clients have different features and no aligned sample ID, such methods cannot establish a reliable bridge. So, this work proposes a personalized federated transfer learning (PFTL) framework, which allows target clients to get the knowledge transferred from the source client but still keep their unique predictive performance.

2.3.2 Recent works on feature heterogeneity in FL

Several recent works have attempted to extend FL beyond homogeneous feature spaces. Rakotomamonjy et al. [90] proposed FLIC, which learns local embedding functions to map client input features into a shared latent space and aligns them with shared class-conditional anchor distributions via Wasserstein distance. The anchors are Gaussian distributions parameterized and learned by neural networks. FLIC achieved a $+3\%$ accuracy gain over baseline FL methods on a real-world Brain-Computer Interfaces dataset with heterogeneous electrode configurations. However, FLIC assumes that all clients can map their observed features into a common latent space and align them via anchors, which is unsuitable for the case since certain predictive features are absent and cannot be embedded. Shi and Kontar [111] developed a personalized FL via domain adaptation (PFL-DA) framework to address both covariate shift (differences in input feature distributions) and concept shift (differences in input-output relationships) across clients. PFL-DA uses a bi-level optimization: (1) local encoders map client-specific inputs into a shared latent space, and (2) personalized decoders make predictions for each client. Their distributed 3D printing case study showed that PFL-DA reduces test loss from 0.25 to 0.23. However, PFL-DA assumes all clients observe sufficient feature information to construct latent representations for classification or prediction, even if from different domains, which is not feasible in the case. Yang et al. [112] proposed a Federated Multi-task Bayesian Network Learning (FMT-NBN) model to address the feature heterogeneity, where clients have both overlapping and distinct features. They adopt a nonparametric additive structural causal model to capture nonlinear dependencies through probabilistic graphs and design a two-step FL algorithm: (1) local clients train a Bayesian Network on their local data, and (2) the central server estimates global task similarities and updates shared parameters iteratively. In their three-phase flow facility (oil-water-gas), FMT-NBN uncovered causal structures closer to the ground truth (local baselines). However, FMT-NBN assumes that each client fully observes its own features for causal structure exploration, which does not hold in the scenario where clients lack predictive or explainable features. Oishi et al. [113] proposed a two-stage model incorporating feature imputation into FL to address the feature missing problem. Each client first trains a local imputation model and exchanges it with others to infer its missing features, after which standard FL is applied via FedAvg. In the UCI Iris dataset, their method improved classification accuracy from 0.899 to 1.000. However, this method assumes that missing features can be reliably inferred from existing ones and is restricted to two-client scenarios.

Beyond the above FL methods, such as domain adaptation and imputation-based methods [114, 115], many related studies that rely on alignment methods have the potential to bridge feature heterogeneity in FL: (1) Generative models (e.g., GANs) have been used to synthesize missing features/data [116, 117]. They require sufficient local data to train reliable generative models, which are often unrealistic in manufacturing, where fatigue data are very limited; (2) Canonical correlation [118] or eigenvector-based methods [119, 120] align latent spaces by matching covariance structures or principal components. These methods implicitly assume that latent feature spaces are structured and comparable across clients. Covariance alignment may align noise or irrelevant structures in disjoint feature settings, distorting predictive mappings; (3) Statistical alignment measures such as KL-divergence or maximum mean discrepancy (MMD) align marginal feature distributions across clients [121]. However, they perform global alignment, which can collapse distinct conditional relationships with the response. Nonetheless, it is a good way to minimize discrepancies among clients. Optimal transport provides a promising alternative by allowing soft, sample-level alignment [122].

2.3.3 Optimal transport in ML and FL

Optimal transport (OT) offers a mathematical framework for comparing and aligning probability distributions to estimate an optimal mapping by minimizing cost functions [122,

Table 2.1: Comparison of FL categories, representative methods, core ideas, and limitations. ✦ indicates clients with different features, but still sufficient predictive information.

Method	FL Type	Heterogeneous features	No shared IDs	Personalization
FedAvg (2017) [123]	HFL	✗	✓	✗
FedProx (2018) [95]	HFL	✗	✓	✗
FedOT (2022) [124]	HFL	✗	✓	✓
FedOTP (2024) [125]	HFL	✗	✓	✓
FedAli (2025) [126]	HFL	✗	✓	✓
splitVFL (2020) [103]	VFL	✓	✗	✗
aggVFL (2024) [91]	VFL	✓	✗	✗
SFTL (2020) [107]	FTL	✦	✗	✗
SFHTL (2022) [108]	FTL	✦	✗	✗
PrADA (2022) [109]	FTL	✦	✗	✗
Oishi (2023) [113]	FTL	✦	✓	✗
PFL-DA (2023) [111]	FTL	✦	✓	✓
FLIC (2024) [90]	FTL	✦	✓	✓
FMT-NBN (2025) [112]	Others	✦	✓	✓
FedCOT (Ours)	**PFTL**	✓	✓	✓

127]. It is from a classical problem of finding the most cost-efficient way to move mass from one probability distribution to another. Formally, given two probability measures μ and ν on a metric space, OT finds a transport plan γ that minimizes the total cost of moving mass between them under a cost function $c(x, y)$ [128]. The resulting OT distance (i.e., Wasserstein distance) provides a geometry-aware measure of discrepancy between distributions, in contrast to distribution-based distance (e.g., KL), moment-based distance (e.g., MMD), or adversarial losses, which may overlook structural correspondences. Computationally, entropic regularization has made OT scalable [129, 130], enabling its wide adoption in modern ML.

In the ML community, OT has been widely applied to domain adaptation, a subfield of transfer learning that trains a model on one data distribution in the source domain and applies it to a similar but different data distribution in the target domain [121, 131]. The joint distribution optimal transport (JDOT) framework [122] introduced label information to guide alignment, ensuring that source and target domains remain predictive with respect to responses. Extensions such as DeepJDOT [132] integrated OT with deep neural networks to improve representation learning in unsupervised adaptation. These works demonstrated that OT-based methods often outperform adversarial or MMD-based alignment methods, as they preserve sample-level correspondences across domains.

Recently, OT has been explored to mitigate statistical heterogeneity in federated settings. Farnia et al. [124] developed a personalized federated learning method based on optimal transport (FedOT). FedOT learns client-personalized transport maps to align local data distributions into a shared global domain, updated iteratively via multi-marginal OT barycenters. FedOT consistently outperformed the baseline method FedAvg with over 15% accuracy improvement on CIFAR-10 and Colored-MNIST under affine and color-based shifts. Li et al. [125] introduced a Federated Prompts Cooperation via Optimal Transport (FedOTP) model to address data heterogeneity in vision-language models by leveraging global and local prompts with unbalanced OT. Each client learns a shared global prompt to create common knowledge and a personalized local prompt to retain client category traits. Unbalanced OT is applied to align local visual features with textual prompts, enabling prompts to focus on the relevant image patches. FedOTP achieves up to 3.7% accuracy gain under label and feature shifts in CIFAR-100 and Office-Caltech10 datasets. Ek et al. [126] developed a personalized FL method - Federated Alignment (FedAli) to mitigate client drift (heterogeneous data). FedAli designed the Alignment with Prototypes layer, which utilizes OT to align input embeddings with global prototypes (i.e., shared knowledge) and local prototypes (client memory representations). It reduces training divergence while enhancing generalization. FedAli outperformed baselines such as FedAvg and FedProx in their heterogeneous Human Activity Recognition datasets and vision benchmarks (CIFAR-10/100).

Despite these advances, existing approaches focus primarily on data distribution alignment without considering semantic meaning. Specifically, they do not incorporate the conditional relationship between features and response, which is critical in both regression and classification tasks. It may cause distorted alignment, leading to poor generalization.

Chapter 3

Defect Characterization and Criticality Analysis on Fatigue Life of Additive Manufactured Parts via Machine Learning

Defects innate in additively manufactured components may lead to inferior and more variation in fatigue lives, thus challenging the qualification of these components in fatigue-critical applications. This work seeks to correlate geometrical features of critical defects measured from fracture surfaces to the fatigue performance of laser beam powder bed fusion (L-PBF) components with machine learning and to develop an integrated data-driven analytical framework for defect criticality (IDADC). IDADC has the potential to enhance the understanding of defect-fatigue relationships in a data-driven fashion. The case study 1 results show that the obtained relationships between the extracted size-related and morphology-related defect features and the fatigue life of L-PBF specimens align with the known fatigue mechanisms and influencing factors. Furthermore, the proposed IDADC framework can model the relationships between defect features and fatigue life with a low mean absolute percentage error of 0.101 using a kernel support vector regression. Additionally, this chapter also proposes a critical direction curvature feature of volumetric defects for nondestructive fatigue life prediction in laser powder bed fusion (L-PBF) Ti-6Al-4V parts via data-driven models. This feature is defined as the maximum curvature component of the defect surface along the loading direction and is extracted from X-ray computed tomography to capture the local geometric sharpness with respect to the loading direction. The case study 2 results show that incorporating this feature substantially improved fatigue life prediction accuracy by 22.46% in linear regression and 35.34% in neural network after removing stress amplitude effect, compared to the baseline with only conventional defect size and location features. This work could establish the algorithmic foundation for nondestructive fatigue evaluation of additive manufacturing products from various facets of critical defects.

3.1 Introduction

Additive manufacturing (AM), which creates intricate geometrical 3D objects using a layer-wise strategy directly from digital solid models, has transformed from rapid prototyping to a revolutionary digital manufacturing technology for functional and structural applications [133–137]. With its superiority in design flexibility, product customization, and low-volume timely production, AM constitutes a viable alternative to traditional methods of manufacturing [135, 136, 138, 139]. As a metal AM process, laser powder bed fusion (L-PBF) melts metallic powders with a heat source in a layer-wise manner during fabrication and enables nearly limitless flexibility in manufacturing. It dominates the current metal AM market with 86.5% of globally installed AM units being L-PBF [140]. One of the key challenges in adopting AM technologies for functional and structural applications lies in the uncertainty of their fatigue performance [141]. Degraded fatigue strength and a pronounced scatter in the fatigue life of L-PBF components can be attributed to the abundance and variability of L-PBF process-induced defects [20]. In L-PBF, process characteristics, such as a highly dynamic melt pool, ultra-high solidification/cooling rates, and a large thermal gradient, inevitably affect the microstructural traits and result in varying levels of process-induced defects, including entrapped gas-entrapped pores, lack of fusions (LOFs), and keyholes [142]. These defects are one of the major life-limiting factors under cyclic loadings and the dominant mechanisms for fatigue crack initiation in L-PBF components. Fatigue performance of L-PBF components, specifically after surface treatments ridding any significant surface roughness, is potentially influenced by the presence of defects that act as a source of stress concentrations [143, 144]. Locations with a large population of more severe defects, such as the LOF defects, may initiate fatigue cracks earlier due to high-stress concentrations in those regions. Research on the impact of defects on the fatigue performance of L-PBF specimens has gained great traction [12–16, 145–148]. Defect features, such as type, size, aspect ratio, and distance to the surface, are extracted from fractography and X-ray computed tomography (CT) images, and incorporated into existing fatigue theories (e.g., fracture mechanics-based fatigue life prediction) for analysis. With the advancement of measurements and data analytics, there is a burgeoning need to leverage computer vision and machine learning (ML) [149] to extract defect features and evaluate the critical roles of defects in the fatigue performance of L-PBF specimens. This study attempts to advance the understanding of defect criticality on fatigue performance by incorporating computer vision and ML in an integrated data-driven analytical framework for defect criticality (IDADC).

3.2 Integrated Data-driven Analytical Framework for Defect Criticality

In this study, an IDADC is developed to explore and understand the relationships between various defect features captured from fractography images and the fatigue life of L-PBF specimens. The fractography images are captured by SEM on the fracture surface of failed fatigue specimens to determine the defect responsible for the crack initiation and failure of the specimens. Only those defects that were identified as being responsible for initiating fatigue cracks were considered in the IDADC. The IDADC consists of three primary elements as follows and is summarized in Figure 3.1:

- 1) Defect feature extraction: content-based image retrieval (CBIR) [150], a computer vision technique, describes defect characteristics on fractography images with various numerical descriptors.

- 2) Correlation analysis: Pearson correlation coefficient (PCC) and maximal information coefficient (MIC) are integrated to analyze the linear and nonlinear relationships between the defect features and fatigue life of L-PBF specimens.

- 3) Defect-fatigue modeling: a novel machine learning model, kernel support vector regression (SVR) with model-agnostic interpretation, is developed to identify the critical defect features and quantify their marginal impacts on the fatigue life of the L-PBF specimens.

This IDADC provides a novel data-driven perspective to investigate the effect of defects on fatigue performance and identify the critical defects that may initiate cracks in the specimen under fatigue loading. It can bridge the current research gaps by examining the defect

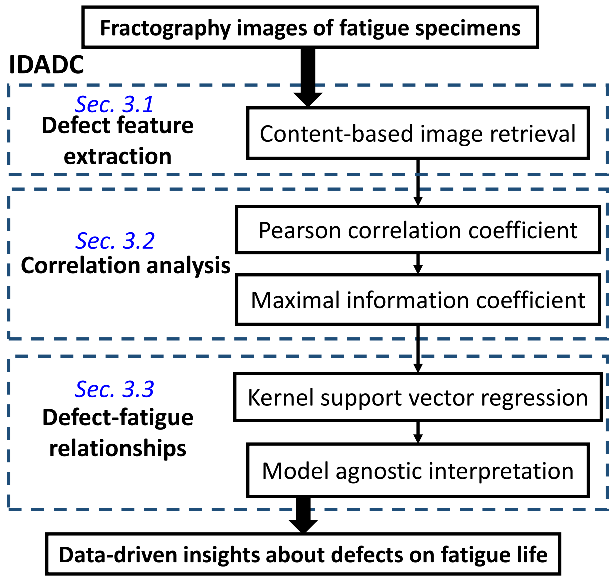

Figure 3.1: The proposed IDADC for understanding defect-fatigue relationships in L-PBF specimens via data-driven analytics.

characteristics systematically and their relationships with the fatigue life of L-PBF specimens. The outcomes and insights from the framework can be used to enhance the understanding of experimental studies and improve fatigue life evaluation for L-PBF materials.

3.2.1 Defect feature extraction

2D defect feature extraction from fractography

For crack-initiating defects (referred to as just defects from hereon) in L-PBF specimens on the 2D fractography images, as shown in Figure 3.2 (a), a variety of defect features can be extracted from the defect contours (some of which are shown in Figure 3.2 (b)) by CBIR to characterize distinctive aspects of defects. In addition to the conventional features, like area, perimeter, circularity, and aspect ratio, other "unconventional" features can also shed light on various defect characteristics. They include: eccentricity, which is the ratio of the distance between the foci of the outer ellipse and its circularity, describes how close the defects are to a circle; solidity, which is the ratio of a defect area filling its convex hull, indicates whether the defects are convex and compact with an irregular shape; and angularity, which calculates the error of a defect contour to fit an ellipse, implies the smoothness of defect boundary and the potential existence of sharp angles. Some of the extracted features are depicted in Figure 3.2 (c).

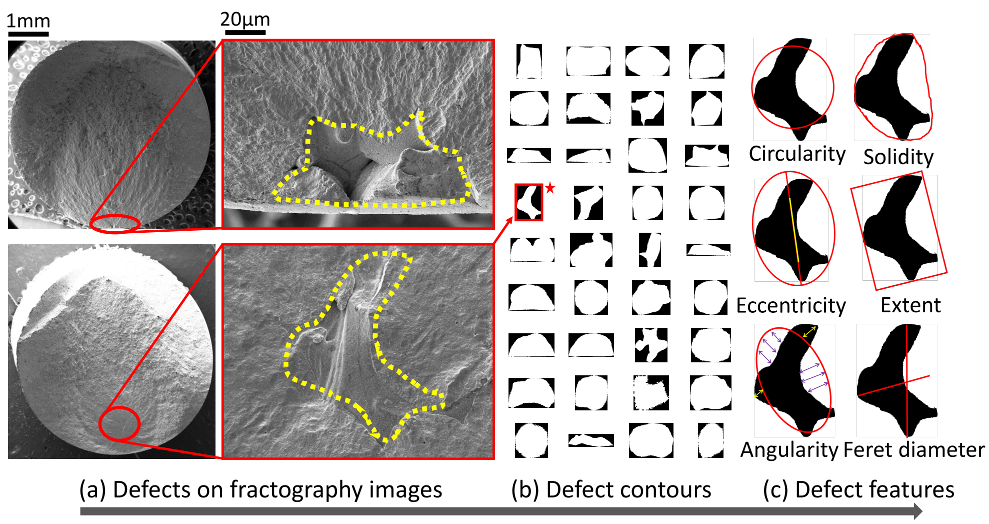

Figure 3.2: The procedure and illustrations of defect feature extraction from fractography images using CBIR techniques.

All the extracted defect features, including by CBIR, are summarized in Table 3.1 with their respective definitions. They can be categorized into size-related, morphology-related, and distance-related features. Moreover, fatigue life is also influenced by subsequent post-processing thermal heat treatments [151, 152] and stress response levels in the fatigue tests. Thus, the strain amplitude and normalized stress level parameter, derived from the stress response and ultimate tensile strength, are also included in the analysis as process-related features. All these features are utilized to quantify various aspects of defects and to investigate the relationships between defects and fatigue life of L-PBF specimens, as they are affected by the level of ductility induced by the post-fabrication heat treatment in this study.

3D defect feature extraction from XCT scans

This section introduces the curvature-based defect feature for nondestructive fatigue life prediction in L-PBF [153]. This work develops a new critical directional curvature feature (from 3D XCT defects), capturing local geometric sharpness considering the loading direction.

Table 3.1: Defect features from fractography images and process-related features of L-PBF specimens.

Category	Descriptors (Features)	Definition
Size	Major axis length	a
	Minor axis length	b
	Area	S
	Convex area	S_C
	Bounding box area	$S_B = a \times b$
	Equivalent diameter	D_E
	Perimeter	C
	Max Feret diameter	D_f^{Max}
	Min Feret diameter	D_f^{Min}
Morphology	Eccentricity	$Eccentricity = \frac{\text{Distance between the foci of the outer ellipse}}{a}$
	Circularity	$Circularity = \frac{4\pi \times S}{S_C^2}$
	Solidity	$Solidity = \frac{S}{S_C}$
	Extent	$Extent = \frac{S}{S_B}$
	Angularity	Error of a shape to fit an ellipse
Distance	Distance to the surface	Distance to the specimen external surface
Process	Strain amplitude	S_a
	Ultimate tensile strength	UTS
	Normalized stress level	$\frac{S_a}{UTS}$

Curvature quantifies the local bending of a surface and is directly related to the local notch severity of a defect. It quantifies local morphology in both 3D XCT and 2D fractography critical defects. At a given surface point, the local geometry is illustrated by two principal directions $\vec{t_1}$ and $\vec{t_2}$ with principal curvatures k_1 and k_2, respectively, as shown in Figure 3.3. These principal curvatures are the maximum and minimum curvatures.

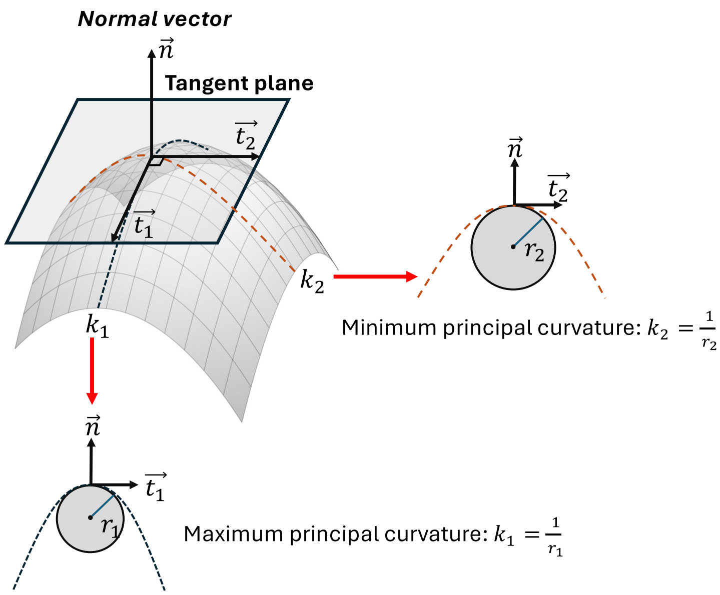

Figure 3.3: Illustration of principal curvatures on a surface. The maximum (k_1) and minimum (k_2) principal curvatures are defined along orthogonal directions $\vec{t_1}$ and $\vec{t_2}$ at a surface point, respectively. The local cross-sectional profiles along each direction are shown, highlighting the different geometric sharpness captured by k_1 and k_2.

To consider the interaction between the local defect curvature and the applied loading, this work proposes the directional curvature, capturing the effective curvature in the loading direction. This work defines a load-consistent tangent vector $\vec{t_L}$, which has the same direction as the projected loading vector $\vec{t_L'}$ of the loading direction $\vec{L}$ on the tangent plane, illustrated in Figure 3.4a. This vector can be expressed as

$$\vec{t_L} = \cos\theta\,\vec{t_1} + \sin\theta\,\vec{t_2} \tag{3.1}$$

where θ is the angle between directional curvature vector $\vec{t_L}$ and maximum curvature vector $\vec{t_1}$

in the tangent plane.

The directional curvature along $\vec{t_L}$ is then calculated based on Euler's equation [154] as

$$k_L = \vec{t_L}^\top \mathbf{S} \vec{t_L} = k_1 \cos^2\theta + k_2 \sin^2\theta \tag{3.2}$$

where **S** is the curvature tensor, or shape operator, defined as

$$\mathbf{S} = \begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix} \tag{3.3}$$

in the $(\vec{t_1}, \vec{t_2})$ basis.

The direction curvature is a weighted combination of the two principal curvatures that satisfies

$$\min(k_1, k_2) \leq k_L \leq \max(k_1, k_2) \tag{3.4}$$

k_L measures the effective curvature aligned with the applied loading direction, governing the local stress concentration along the loading direction. It has a direct geometric-mechanical link with classical notch theory, which relates the stress concentration factor (SCF) to local curvature. For an ideal notch (e.g., ellipse) with local radius of curvature r, the SCF is defined as

$$K_t = 1 + 2\sqrt{\frac{a}{r}} \tag{3.5}$$

where a is the notch depth. It indicates sharper corners in defects with higher curvature, which cause larger stress concentrations and earlier crack initiation.

According to Figure 3.4b, to identify the high-stress concentration zone, this work considers the convex surface ($k_1 > 0, k_2 > 0$) in 3D XCT critical defects, as these geometries amplify the applied stress through local curvature effects.

Figure 3.5 demonstrates four representative cases of 3D XCT defects based on the signs of the principal curvatures:

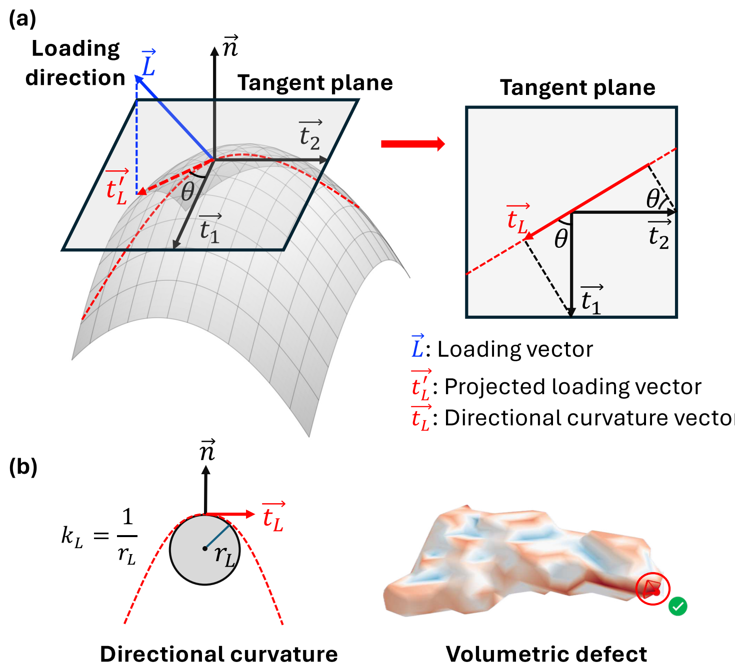

Figure 3.4: Illustration of the defined directional curvature on a 3D surface. (a) Derivation of the direction curvature vector $\vec{t}_L$ based on the loading vector $\vec{L}$ and the principal curvature vectors $\vec{t}_1$ and $\vec{t}_2$. (b) Demonstration of directional curvature and its corresponding curvature maps with a marked effective stress concentration region in a critical volumetric defect.

1. Convex dome ($k_1 > 0, k_2 > 0$): the surface bulges outward relative to the material, becoming a potential crack initiation point due to stress concentration.

2. Saddle ($k_1 > 0, k_2 < 0$): the surface has mixed curvatures with opposing bending directions, leading to less stress concentration.

3. Concave bowl ($k_1 < 0, k_2 < 0$): the surface curves inward relative to the material, experiencing stress relaxation rather than concentration.

4. Cylinder ($k_1 < 0, k_2 = 0$): minimal stress concentration.

For each reconstructed internal defect mesh surface, the directional curvature is evaluated at every surface vertex via Eq. (3.2). Local principal curvatures k_1, k_2 and direction vectors $\vec{t}_1, \vec{t}_2$ are captured by quadratic surface fitting.

To reduce numerical noise and balance the resolution issue, the surface is smoothed with

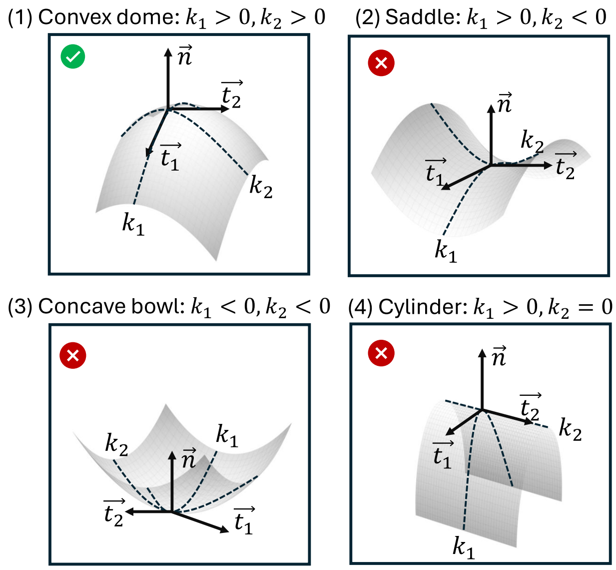

Figure 3.5: Categories of local surface geometries with respect to principal curvatures k_1 and k_2. Only the convex dome condition ($k_1 > 0, k_2 > 0$) concentrates stress, where the convex geometry acts as a notch and bulges into the material under loading.

Gaussian noise before computing its curvature. The critical directional curvature is defined as the absolute maximum curvature in the convex region ($k_1 > 0, k_2 > 0$):

$$\text{Critical directional curvature} = \max_{x \in \Omega} |k_L(x)| \tag{3.6}$$

where Ω is the convex subset of the defect surface and x represents a point on the 3D defect surface. This feature denotes the sharpest convex geometry aligned with loading, indicating the potential site for stress concentration and crack initiation.

3.2.2 Correlation analysis based on Pearson correlation coefficient and maximal information coefficient

In the proposed IDADC, the Pearson correlation coefficient (PCC) [155] and maximal information coefficient (MIC) [156] are utilized for correlation analysis to explore and quantify the linear/nonlinear relationships between the fatigue life of L-PBF specimens and various defect features, as well as the interdependence among those defect features. PCC is a widely used statistic to measure the linear correlation between two variables X and Y.

$$r_{XY} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{3.7}$$

where $\{(x_i, y_i), i = 1, ..., n\}$ are the observed data points for the two variables, n is the number of data points in the dataset. PCC takes values between -1 and 1, with 0 indicating statistical independence between variables X and Y, -1 indicating a completely negative relationship, and 1 indicating a completely positive relationship.

Moreover, MIC is an information-theory-based measure of association that can capture a wide range of linear or nonlinear relationships between variables. MIC is formulated based on a naive mutual information estimate $I_{\text{MIC}}\{x, y\}$ via a data-dependent binning scheme:

$$\text{MIC}(x, y) = \max\left(\frac{I_{\text{MIC}}\{x, y\}}{\log_2 \min\{n_X, n_Y\}}\right) \tag{3.8}$$

$$I_{\text{MIC}}\{x, y\} = \sum_{\tilde{x},\tilde{y}} \hat{p}(\tilde{x}, \tilde{y}) \log_2 \frac{\hat{p}(\tilde{x}, \tilde{y})}{\hat{p}(\tilde{x})\hat{p}(\tilde{y})} \tag{3.9}$$

where n_X, n_Y denote the number of bins imposed on the x,y axes; $\hat{p}(\tilde{x}, \tilde{y})$ is the fraction of data points falling into the bin $(\tilde{x}, \tilde{y})$. The MIC binning scheme is chosen to maximize the ratio in Eq. (3.8) with the total number of bins $(n_X \times n_Y)$ in the user-specified threshold. MIC takes values between 0 and 1, where 0 indicates statistical independence between variables X and Y, and 1 indicates complete noiseless dependence. However, MIC is computationally expensive and does not indicate the direction or the sign of the relationships.

Correlation analysis is a straightforward and intuitive data-driven step to understand the relationships between defect features and fatigue life. The results from correlation analysis can highlight the critical defect features and the strength/direction they impact the fatigue life.

3.2.3 Defect-fatigue relationship analysis based on kernel support vector regression

The correlation analysis in Section 3.2 explores the relationships between the fatigue life of L-PBF specimens and individual defect features. The IDADC then utilizes the kernel support vector regression (SVR) to understand how defect features concurrently impact fatigue life by establishing a surrogate model with model-agnostic interpretation.

In IDADC, SVR [157] aims to formulate a nonlinear surrogate model using a kernel function to predict the fatigue life of an L-PBF specimen based on various defect features. It can map the nonlinear relationships between defect features and fatigue life into a high-dimensional space (which can be linearly segmented) via the kernel function and approximate the fatigue life with linear relationships in this high-dimensional space.

With the defect features of L-PBF specimens extracted by CBIR in Section 3.1 and the experimentally obtained fatigue life values, this work denotes the dataset as $D_i = (\mathbf{x}_i, y_i), i = 1, 2, 3, ..., n$, where $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{id})^T$ are d defect features of the specimen i, and y_i denotes the fatigue life of the specimen i. Radial basis kernel $K(\mathbf{x}_i, \mathbf{x}_j) := \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle = e^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2}$ has a strong ability to fit nonlinear and high-dimensional defects data sets and it overcomes the space complexity problem as it just needs to store the support vectors. Therefore, a nonlinear mapping function $\Phi(\mathbf{x}_i)$ is implicitly represented by a radial basis kernel used for all data sets in this SVR, with $\gamma > 0$ (it controls the curvature of the decision boundary) selected by grid search. The SVR model can be simply written in a matrix format

$$f(\Phi(\mathbf{X})) = \Phi(\mathbf{X})\mathbf{w} + b \tag{3.10}$$

where $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)^T$ is the matrix for defect features in the training data, $\mathbf{w} = (w_1, w_2, ..., w_d)^T$ are the regression coefficients, and b is the bias in regression. To solve this SVR problem, its objective function can be formalized as follows

$$\begin{aligned}
\min_{\mathbf{w},b} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\left(\xi_i + \hat{\xi}_i\right) \\
\text{s.t.} \quad & \Phi(\mathbf{x}_i)\mathbf{w}^T + b - y_i \leq \epsilon + \xi_i, \\
& y_i - \Phi(\mathbf{x}_i)\mathbf{w}^T - b \leq \epsilon + \hat{\xi}_i, \\
& \xi_i > 0,\ \hat{\xi}_i > 0, \quad i = 1, 2, \dots, n.
\end{aligned} \tag{3.11}$$

where $\|w\|^2$ is the magnitude of the normal vector to the surface that is being approximated, C denotes the regularization constant, ξ_i and $\hat{\xi}_i$ are slack variables to guard against outliers, and ϵ is the soft margin in SVR. This model can be solved by the Lagrange multiplier method while satisfying the Karush–Kuhn–Tucker (KKT) conditions [158, 159]. Consequently, with the SVR in Eq. (3.11), the relationships between defect features and fatigue life of L-PBF specimens can be modeled in a data-driven fashion.

3.2.4 Interpretation with model-agnostic methods

In IDADC, model-agnostic interpretation methods are integrated to explain the importance of defect features in determining the response of fatigue life in "black-box" SVR, as illustrated through a variety of graphics generated based on importance analysis [160]. Permutation feature importance (PFI) and accumulated local effects (ALE) are used in this study. PFI measures the importance of a feature by calculating the increase in the model error after permuting the feature. A feature is "important" if shuffling its values increases the model error. Mean absolute error (MAE) is selected as the error function in this study.

ALE describes the dependence of features on ML outcomes, even when the features are correlated. It demonstrates the marginal effect of interested features (in set S) on the ML outcomes by marginalizing the ML model output over the distribution of the other features in set C. Therefore, the function shows the relationships between the interested features and the model outcomes. It is realized by averaging the changes in the model outcomes and accumulating them over the grid. To estimate local effects, this work divides the feature into many intervals and computes the differences in the model outcomes.

$$\hat{f}'_{S,\mathrm{ALE}}(x) = \sum_{k=1}^{k_S(x)} \frac{1}{n_S(k)} \sum_{i:x_S^{(i)} \in I_S(k)} \left[f\left(z_{k,s}, \mathbf{x}_C^{(i)}\right) - f\left(z_{k-1,s}, \mathbf{x}_C^{(i)}\right) \right] \tag{3.12}$$

where $\hat{f}$ is the ML surrogate model; $\mathbf{x}_S$ are the interested features in set S and $\mathbf{x}_C$ are the other features in set C; z's denote a grid of intervals over which this work computes the changes in the outcome of the ML model. This effect is centered so that the mean effect is zero.

$$\hat{f}_{S,\mathrm{ALE}}(x) = \hat{f}'_{S,\mathrm{ALE}}(x) - \frac{1}{n} \sum_{i=1}^{n} \hat{f}'_{S,\mathrm{ALE}}\left(\mathbf{x}_S^{(i)}\right) \tag{3.13}$$

ALE plots are centered at zero, and the value at each point of the ALE curve is the difference from the mean outcomes. Thus, they indicate the relative effect of changing the feature on the model outcomes.

3.3 Experimental Setup

3.3.1 Case study 1: L-PBF 17-4 PH SS fatigue test specimens

Materials, fatigue testing, XCT and fractography

Fracture surface information from L-PBF 17-4 PH SS fatigue test specimens is used for integrated data-driven analysis and modeling in this study. The samples were fabricated as part of three previous studies [145, 161, 162] on an EOS M290 L-PBF AM system under an argon-shielded atmosphere [145]. EOS recommended process parameters for 17-4 PH SS, presented in Table 3.2, were used to fabricate 11 mm square bars in studies [161, 162] and three different geometries of dog-bone, small block, and large block in study [145]. Subsequently, all the printed samples were surface machined to fatigue specimen geometries following recommendations from ASTM E606, details of which can be found in [145, 161, 162]. Heat-treatment was performed for the different batches of specimens following ASTM A693 [57] using five conditions of H900, H1025, CA-H900, CA-H1025, and CA-H1150). Fatigue tests were performed in the previous studies according to ASTM E606 under fully-reversed (R = -1), strain-controlled mode under strain amplitudes of 0.0025 mm/mm, 0.003 mm/mm, and 0.004 mm/mm on an MTS servo-hydraulic load frame [145, 161, 162]. The stress response of the material was recorded from the fatigue tests as well as the strain-life data. Tensile tests were also performed as per ASTM E8 to determine some of the parameters used in the analysis, such as the yield strength (YS) and the ultimate tensile strength (UTS). A combination of optical microscopy using a Keyence VHX-6000 capable of imaging up to a 1,000x magnification and scanning electron microscopy using a Zeiss Crossbeam 550 SEM capable of imaging up to a 1,000,000x magnification, was performed to acquire the fractography images. The resolution for these two imaging techniques depends on the scan settings, and can be up to 1 μm for optical microscopy and 4 nm for scanning electron microscopy. More details on the fabrication, post-processing, tensile, and fatigue testing conditions can be found in references [145, 161, 162]. The final image data was split into categories based on the testing strain amplitudes and the heat treatments for training/validating the developed IDADC framework.

Table 3.2: Default L-PBF core process parameters for 17-4 PH SS as provided by EOS and used to fabricate all the specimens in this study.

Laser power (W)	Scanning speed (mm/s)	Hatching distance (μm)	Layer thickness (μm)
220	755.5	100	40

Data acquisition

One hundred and ninety 2D fractography images of critical defects for L-PBF 17-4 PH SS specimens (in machined surface condition) are captured. Critical defects on the fracture surface are the defects likely to initiate the cracks. If there are multiple defects on the fracture surface and they are far away from each other, this work identified one critical defect based on experience and extracted its features for the analysis. If defects are connected on the fracture surface, this work treats them as one critical defect. There are six main categories based on the different strain amplitudes in the strain-controlled fatigue tests (i.e., 0.001 mm/mm, 0.0015 mm/mm, 0.0020 mm/mm, 0.0025 mm/mm, 0.003 mm/mm, 0.004 mm/mm) with 8, 6, 47, 17, 62, and 50 defects, respectively. It is worth noting that the dataset for the L-PBF specimens under different strain-controlled fatigue tests and heat treatments used in this study is limited for the proposed ML methods. In general, the model accuracy of ML relies on both the data size and the data quality (e.g., the consistent conditions under which the datasets are collected) [149]. Therefore, this work selected three disparate datasets for analysis and investigation in this study by making tradeoffs between the data size and consistency in conditions (i.e., strain-controlled fatigue tests and heat treatments). Any outliers were removed from the selected datasets based on the key values of fatigue life, defect size, and the distance proximity to the surface shown in Figure 3.6. The thresholds for separating outliers from all data are 330,984 cycles, 143,400 μm^2 and 6.3 μm for fatigue life, defect size, and the distance proximity to the surface, respectively.

1. Dataset 1 (large data size, but low consistency): 157 defects on all L-PBF 17-4 PH SS fatigue testing specimens.

2. Dataset 2 (medium data size, and medium consistency): 51 defects on L-PBF 17-4 PH SS specimens with different heat treatments in fatigue tests at a strain amplitude of 0.003 mm/mm.

3. Dataset 3 (small data size, but high consistency): 34 defects on L-PBF 17-4 PH SS specimens with heat treatment of CA-H1025 and fatigue tests at a strain amplitude of 0.003 mm/mm.

Table 3.3: Datasets used in the IDADC for understanding defect-fatigue relationships.

	Strain level (mm/mm)	Heat treatment	Total data points	Outliers removed	Remaining data points
Dataset 1	All	All	190	33	**157**
Dataset 2	0.003	All	62	11	**51**
Dataset 3	0.003	CA-H1025	43	9	**34**

The experimental datasets used to understand the relationships between defect features and fatigue life are presented in Table 3.3. In addition, this work will investigate whether data size or consistency in conditions is important in understanding the relationships in the proposed IDADC framework.

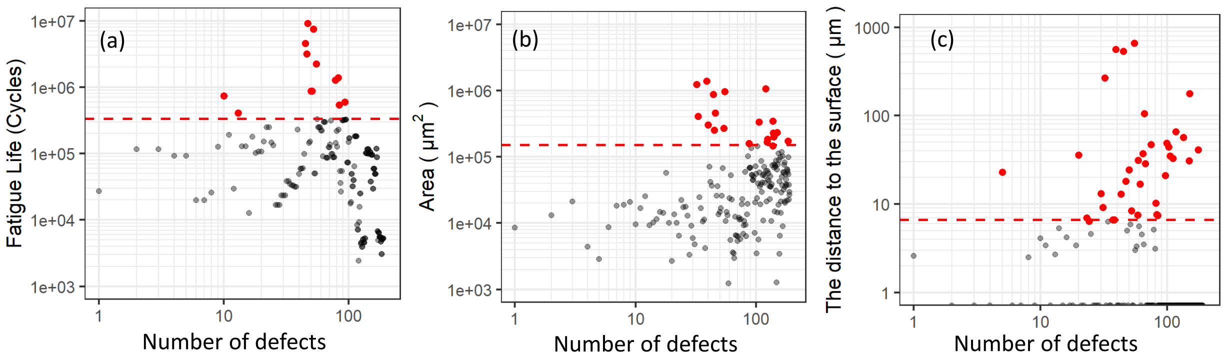

Figure 3.6: The selection of outliers for all data. (a) Outliers of fatigue life of L-PBF specimen; (b) Outliers of area of L-PBF specimen; and (c) Outliers of the distance to the surface of L-PBF specimen. Note: Red points mean outliers, red dash line is the threshold between outliers and normal points and all data are displayed in a logarithmic coordinate system)

3.3.2 Case study 2: L-PBF Ti-6Al-4V fatigue test specimens

As shown in Figure 3.7, Ti-6Al-4V cylinders were fabricated using an EOS M290 L-PBF system from the National Center for Additive Manufacturing Excellence (NCAME) at Auburn University. According to Table 3.4, fourteen L-PBF Ti-6Al-4V specimens were produced under six fabrication conditions and two printing orientations, using varied energy densities from 44.44 to 95.24 J/mm^3 to induce overheating, underheating, and associated defects such as KHs and LoFs. After fabrication, these parts were stress-relieved at 705°C for one hour with a heating rate of 5°C per minute in an electric furnace, followed by furnace cooling before removal from the build plate. Furthermore, all specimens were machined and polished to achieve a low surface roughness suitable for uniaxial fatigue testing and defect criticality analysis.

Nondestructive inspection: XCT scanning and defect reconstruction The 12 mm gauge shown in Figure 3.7 was scanned using a Zeiss Xradia 620 Versa XCT machine at 160 kV and 21 W with an X-ray filter to reduce low-energy photons, maintaining a voxel size of approximately 5.5 μm. The volumetric data were reconstructed with Zeiss Xradia software to assess internal defects in the

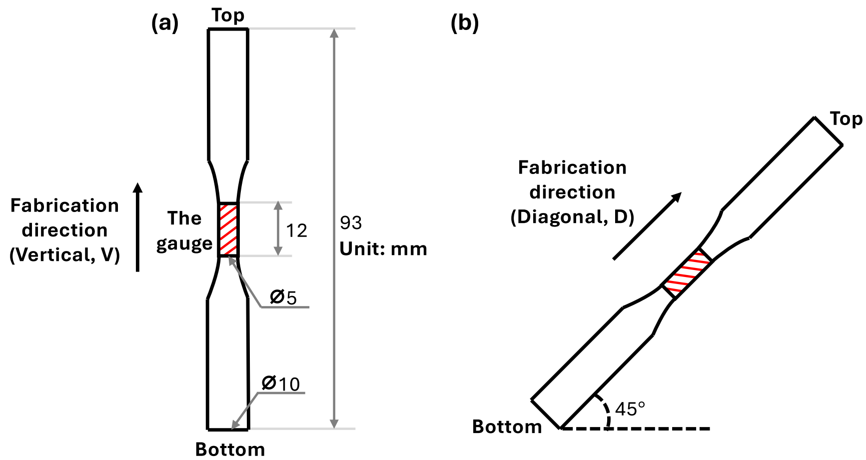

Figure 3.7: The design of geometry and dimensions of the L-PBF fatigue test specimens in two build directions: (a) vertical, and (b) diagonal, from bottom to top.

specimens [163]. Finally, the reconstructed grayscale XCT images were binarized in the ImageJ software for defect segmentation [164].

Identification of critical defects The critical defects in fourteen specimens were identified in XCT scans by leveraging positional features (e.g., distance to the surface, Z-location, and relative position to a designated reference tape), size (e.g., feret diameter), and morphology-related features (e.g., aspect ratio and circularity) obtained from the critical defects observed in SEM fractography. Based on these criteria, candidate defects with similar position, size, and morphology to the critical defect were identified. A dissimilarity (DS) score was then computed for each candidate defect, representing a weighted combination of positional, size-, and morphology-related feature distances between each candidate defect and the critical defect. The candidate defect with the lowest DS was ultimately designated as the critical defect [163]. In this work, all the identified critical defects agree with the corresponding observations from SEM fractography.

Surface meshing and curvature calculation After XCT reconstruction, the 3D XCT image stacks were processed to identify internal defects and extract their morphology features. A Gaussian filter was first applied to reduce pixel noise, followed by Otsu's thresholding method to segment and binarize the defect regions. The binarized volumetric defects were then converted into surface meshes using the Python vedo package, which allows efficient mesh generation and surface

Table 3.4: Fabrication conditions and induced defect types for fourteen laser L-PBF Ti-6Al-4V specimens. The table shows the fabrication directions, process conditions, and specimen design employed to produce keyholes (KHs) and lack-of-fusions (LoFs) [163].

Fabrication Direction	**Process Condition**	**Energy Density** (J/mm^3)	**Laser Power (W)**	**Scanning Velocity** (mm/s)	**Hatch Spacing** (μm)	**Layer Thickness** (μm)	**Specimen ID**
Vertical	LaV (LoF)	50.00	252	1200	140	30	LaV04, LaV15
	LbV (LoF)	44.44	224	1200	140	30	LbV07
	LcV (LoF)	46.29	280	1200	168	30	LcV01, LcV07, LcV10, LcV11
	RV (Recommended)	55.55	280	1200	140	30	RV05
	KaV (KH)	95.24	336	840	140	30	KaV04, KaV11, KaV15
	KbV (KH)	90.28	364	960	140	30	KbV01, KbV02
Diagonal	RD (Recommended)	55.55	280	1200	140	30	RD09

smoothing for defects with complex geometries. Afterward, local curvatures were obtained from the reconstructed 3D meshes based on an area-weighted average of adjacent face normals, a least-squares quadric fit surface, and the shape operator [165].

Feature normalization and stress-adjusted fatigue life Computer vision-based methods [11] can extract geometric features from 3D XCT and 2D fractography critical defects, including size-related features such as Murakami's $\sqrt{\text{area}}$ and maximum axis length, as well as morphology-related features such as aspect ratio and solidity (i.e. the ratio of the defect area/volume to its convex hull area/volume), to characterize defects and relate them to fatigue performance under cyclic loading. All features were normalized to a range of 0 to 1 using the min–max normalization method. To remove the dominant effect of stress amplitude and isolate the influence of defect features, a linear regression model based on Basquin's equation [166, 167] was fitted to obtain the stress-adjusted fatigue life.

Experimental setup and validation strategy

This work employs rigorous validation using Leave-One-Out Cross-Validation (LOOCV) to address the limited number of available fatigue test specimens, ensuring the generalizability and reliability of the proposed approach. Specifically, in each iteration, one test specimen was excluded from the training set and used as a test sample until all samples had been tested once. This method provides a robust and low-bias estimate of model performance, maximizing data utilization.

To systematically evaluate and validate the impact of the proposed curvature features, fatigue life prediction performance, and feature importance, three scenarios were considered:

1. **Baseline case:** stress-adjusted fatigue life prediction based only on defect size and the distance to the surface;

2. **Curvature-enhanced case:** stress-adjusted fatigue life prediction based on defect size and the distance to the surface with additional proposed curvature feature;

3. **Full feature case:** stress-adjusted fatigue life prediction based on defect size, the distance to the surface, proposed curvature feature, and solidity.

This work evaluates model performance for nondestructive and destructive fatigue life prediction via Root Mean Square Error (RMSE), adjusted coefficient of determination (adjusted R^2), and Symmetric Mean Absolute Percentage Error (SMAPE). The adjusted R^2 was used to account for the different number of input features, and SMAPE provided a more balanced and robust measure of prediction accuracy. In addition, this work employs Shapley Additive Explanations (SHAP) [168], a model-agnostic technique in Explainable AI, to identify the most important and influential features, thereby explaining the nonlinear relationship between these features and the predicted stress-adjusted fatigue life. This approach enhances the physics interpretability of the model output.

For both destructive fatigue analysis and nondestructive fatigue life prediction, this work uses grid search to select the optimal architecture and hyperparameters, and trains models until convergence to a stable validation error, thereby preventing overfitting and achieving a balanced bias-variance performance. Specifically, in the nondestructive fatigue life prediction, the optimal neural network architecture is intentionally compact, consisting of one hidden layer with 3 neurons, a sigmoid activation function, and a learning rate of 0.01. The sigmoid activation was selected via grid search due to its smooth and bounded nonlinear mapping (-1 to 1), which is suitable to capture nonlinearity while maintaining stable training behavior of small datasets.

To further explore the physical relevance of 3D curvature-based features, this work considers using Eq. (3.5) to estimate SCF (K_t) on the 3D XCT defect surface via notch theory and finite element analysis (FEA) to calculate the local stress distribution near the 2D fractography critical defect via linear elastic principle. The FEA simulations were performed using Autodesk Fusion. A model-based mesh size was adopted with an average element size on the order of 1-10% of the defect size. Quadratic (parabolic) elements with curved mesh representation were used to accurately capture the defect geometry, particularly in regions with high curvature. Each defect was modeled as a thin surface (i.e., an idealized planar defect) embedded in the center of a cube with a side length at least twenty times the length of the critical defect's major axis.

3.4 Results of Case Study 1: Fatigue Life Prediction and 2D Defect Criticality Analysis of L-PBF Parts

3.4.1 Feature extraction of 2D fractography defects

Defect feature extraction from the 2D fractography images is implemented via CBIR on the three datasets in this study. As shown in Table 3.1, the extracted defect features describe various aspects of defect characteristics. These defect features can be summarized statistically to provide an overview of the defects. For instance, the statistics of defect features extracted from Dataset 3 (strain level of 0.003 mm/mm and heat treatment of CA-H1025) are summarized in Table 3.5. It is noticeable from Table 3.5 that:

1. The distributions of size-related features are skewed to the right, and most defects have smaller values than their means. This indicates that exceedingly large defects are uncom-

mon in the specimens.

2. The morphology-related features are within the range $[0, 1]$, and are not as skewed as size-related features. These numbers describe whether the defects are close to an ellipse or a circle (eccentricity, circularity), whether they have a regular and smooth boundary or sharp angles (solidity, angularity), and whether they are convex and compact (solidity, extent).

3. Around half of the defects are open on the surface (with distance to the surface as $0\,\mu$m).

4. The distributions of the fatigue lives of these specimens are also skewed to the right, with the majority of specimens having shorter fatigue lives than their mean. The longest fatigue life is around six times the mean and ten times the median. This indicates that some scatter exists in the fatigue life data at the strain amplitude of 0.003 mm/mm.

Table 3.5: Statistics summary of CBIR-extracted defect features for Dataset 3 (strain level of 0.003 mm/mm and heat treatment of CA-H1025).

Category	Defect features	Mean	Std.	Min.	Median	Max.
Size	Area (μm^2)	3320.2	9271.2	41.1	450.5	46215.8
	Convex area (μm^2)	3334.5	9501.7	43.2	432.6	49642.5
	Major axis length (μm)	53.1	58.6	7.7	30.3	275.8
	Minor axis length (μm)	32.9	41.4	6.7	19.2	209.8
	Equivalent diameter (μm)	40.0	47.0	6.9	22.7	230.1
	Perimeter (μm)	179.4	217.3	29.9	98.1	1115.2
	Max Feret diameter (μm)	55.7	62.9	8.4	30.9	304.1
	Min Feret diameter (μm)	33.6	42.1	7.3	18.5	216.4
Morphology	Eccentricity	0.7	0.2	0.2	0.7	1.0
	Circularity	0.5	0.1	0.3	0.6	0.8
	Solidity	0.9	0.0	0.8	1.0	1.0
	Extent	0.7	0.1	0.4	0.7	0.9
	Angularity	0.2	0.1	0.0	0.2	0.4
Distance	Distance to the surface (μm)	27.3	93.0	0.0	0.0	562.4
Process	Fatigue life (Cycles)	209015	270613	26735.0	126971.0	1395858.0

These defect features constitute the explanatory variables in the dataset for discovering the relationships between defect features and fatigue life of specimens, while the fatigue life of these L-PBF 17-4 PH SS specimens is the target variable in the following analysis and modeling. Furthermore, since fatigue life is also affected by subsequent post-processing thermal heat treatments [151, 152] and strain amplitude in the fatigue tests, a normalized stress level parameter (derived from UTS and the stress level to adjust the effect of heat treatment) and the strain amplitude are also included as explanatory variables in further analysis.

3.4.2 Correlation analysis

Correlations between defect features and fatigue life

Correlation analysis between various defect features and fatigue life is performed using PCC and MIC for the three fatigue testing datasets presented in Table 5. PCC captures positive/negative linear correlations between fatigue life and respective defect features, while MIC indicates the magnitudes (without signs) of (linear and nonlinear) correlations between fatigue life and respective defect features, as shown in Table 5.

Some insights can be inferred from the results in Table 5 for the three datasets:

1. Dataset 1:

    - Strain amplitude in strain-controlled fatigue tests and normalized stress level (adjusted by the heat treatment) of materials have much larger PCC and MIC values than the defect features.
    - The defect features manifest relatively small PCC with fatigue life; however, their impacts on the fatigue life align with known fatigue mechanisms and influencing factors, e.g., defects with larger values in size-related features lead to shorter fatigue life, specimens with circular and less elongated defects tend to have longer fatigue lives, and defects with a smaller value of the distance to the surface lead to shorter fatigue life.

2. Dataset 2:

- Since these specimens are in the same strain-controlled fatigue tests (strain level of 0.003 mm/mm), the defect features are not dominated by the testing conditions but exhibit strong relationships with fatigue life. In addition, they have the same positive/negative impacts on fatigue life as in Dataset 1.
- Both PCC and MIC identify similar defect features having strong relationships with the fatigue life of L-PBF 17-4 PH SS specimens. They are the size-related features (perimeter, max Feret diameter, major axis length, equivalent diameter, convex area, area, filled area) and morphology-related features (circularity, aspect ratio, extent). The distance to the surface also shows a positive correlation, i.e., defects with a smaller value of the distance to the surface lead to shorter fatigue life.

3. Dataset 3:

   - All specimens underwent the same heat treatment, CA-H1025, and were tested at the same strain amplitude, i.e., 0.003 mm/mm; this dataset could provide the most accurate results due to the consistent conditions in the test specimens.
   - Despite a small number of test specimens, the PCC and MIC identify similar size-related and morphology-related defect features as in Dataset 2, which have strong relationships with the fatigue life of L-PBF specimens.
   - The distance to the surface also shows a similar positive correlation as in Dataset 2, i.e., defects with a smaller value of the distance to the surface lead to shorter fatigue life.

Despite the differences in these three datasets, both PCC and MIC identify some shared size-related and morphology-related defect features with high correlation values with the fatigue lives of the L-PBF 17-5 PH SS specimens, bolded in Table 3.6. For instance, these defect features have high PCC and MIC values in all three datasets: max Feret diameter and major axis length (indicate the size and the elongation of the defects), perimeter (indicates the size of the defects and the smoothness of the defect contours), as well as eccentricity and circularity (indicate the

Table 3.6: PCC and MIC between fatigue life of L-PBF 17-4 PH SS specimens and all the defect features for three datasets in this study. The bold values indicate relatively high correlations (≥ 0.5).

Defect Features	Dataset 1		Dataset 2		Dataset 3	
	PCC	MIC	PCC	MIC	PCC	MIC
Strain amplitude	**0.57 (-)**	**0.85**	N/A	N/A	N/A	N/A
Normalized stress level	**0.53 (-)**	**0.84**	0.11 (-)	0.47	N/A	N/A
Major axis length	0.24 (-)	**0.54**	**0.59 (-)**	**0.69**	**0.57 (-)**	**0.72**
Circularity	0.22 (+)	**0.56**	**0.62 (+)**	**0.64**	**0.52 (+)**	0.46
Max Feret diameter	0.22 (-)	**0.57**	**0.60 (-)**	**0.58**	**0.58 (-)**	**0.62**
Eccentricity	0.21 (-)	**0.54**	**0.52 (-)**	**0.74**	0.43 (-)	**0.56**
Perimeter	0.21 (-)	0.49	**0.61 (-)**	**0.59**	**0.58 (-)**	**0.53**
Solidity	0.18 (+)	0.41	0.49 (+)	0.54	0.25 (+)	0.37
Extent	0.18 (+)	**0.61**	**0.54 (+)**	**0.51**	0.38 (+)	**0.56**
Minor axis length	0.15 (+)	0.33	0.07 (-)	0.37	0.35 (+)	0.40
Min Feret diameter	0.13 (+)	0.34	0.41 (-)	0.37	0.39 (-)	0.34
Distance to the surface	0.13 (+)	0.41	0.35 (+)	0.45	0.18 (+)	0.43
Equivalent diameter	0.12 (-)	0.42	**0.54 (-)**	0.46	**0.52 (-)**	0.41
Convex area	0.08 (-)	0.44	**0.55 (-)**	**0.54**	**0.51 (-)**	0.49
Area	0.08 (-)	0.42	**0.53 (-)**	**0.46**	**0.50 (-)**	0.41
Filled area	0.08 (-)	0.42	0.53 (-)	0.46	**0.50 (-)**	0.41
Angularity	0.05 (+)	0.42	0.07 (-)	0.44	0.07 (+)	0.39

* Sorted by descending order of PCC in Dataset 1.

elongation and the roundness of the defects). Therefore, these individual defect features exhibit the relationships with fatigue life, and together they can characterize different types of defects (e.g., gas-entrapped pores, LOFs) and enhance the understanding of their impacts on fatigue life. For instance, compared to LOFs, gas-entrapped pores are usually smaller (with a small perimeter, a small max Feret diameter, and/or a small major axis length) and rounder (with a large circularity and/or a small eccentricity). If gas-entrapped pores are critical defects of an L-PBF specimen, the fatigue life of the specimen tends to be longer than a counterpart L-PBF specimen with LOFs as the critical defects.

Correlations among features in defect analysis

This work further explores the correlations among the features and visualize their interdependence in correlation matrices in Figure 3.8. These features are categorized into the size-related defect features, morphology-related defect features, the distance to the surface, and process features (normalized stress level and strain amplitude).

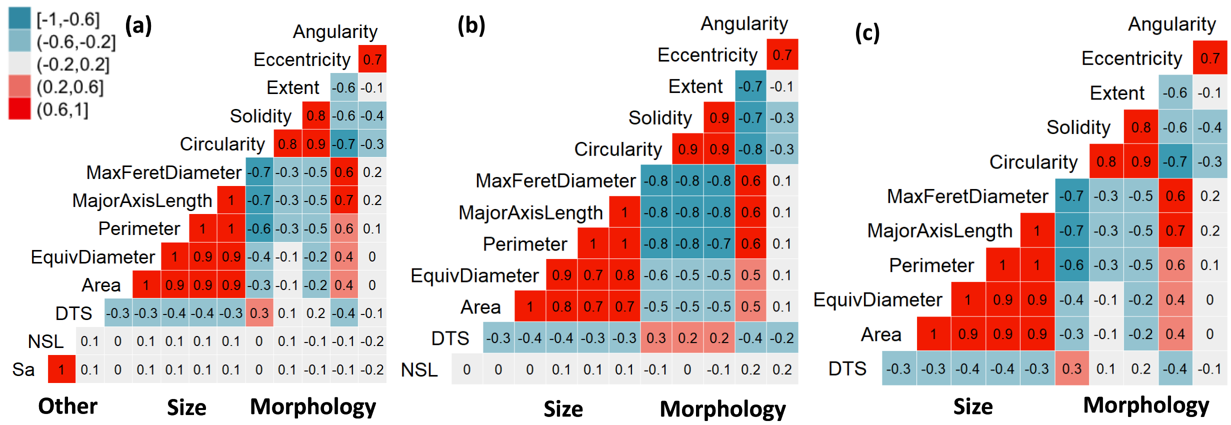

Figure 3.8: PCC among extracted features of L-PBF 17-4 PH SS specimens: (a) Dataset 1 with all specimens; (b) Dataset 2 with fatigue tests at a strain amplitude of 0.003 mm/mm; and (c) Dataset 3 with heat treatment of CA-H1025 and fatigue tests at a strain amplitude of 0.003 mm/mm. (NSL: Normalized stress level; DTS: The distance to the surface)

Within each category of defect features, the correlations among the respective features are relatively high. For example, the size-related features, e.g., area, perimeter, equivalent diameter, major axis length area, are all highly correlated with each other. Between the categories of defect features, it is noted that a few size-related defect features (i.e., max Feret diameter, major axis length, perimeter) have strong correlations with morphology-related defect features (eccentricity, circularity, solid, extent). In detail, defects with a large max Feret diameter, or major axis length, or perimeter tend to have noncircular shapes (with large eccentricity but small circularity, solidity, and extent). These defects are most likely LOFs; therefore, such correlations can help in the classification of defects and the determination of their effects on the fatigue behavior of the materials.

3.4.3 Defect-fatigue data-driven surrogate modeling

Model accuracy

In the IDADC framework, SVR creates a surrogate model to determine the relationships between defect features and fatigue life for the three datasets. The modeling accuracy is evaluated using the mean absolute percentage error (MAPE), which represents the percentage error between the actual fatigue lives and the SVR outcomes, via 5-fold cross-validation.

Three benchmark methods —least absolute shrinkage and selection operator (LASSO), decision tree, and random forest —are adopted in this study to compare with SVR. LASSO fits the linear relationships between the target variable (i.e., fatigue life) and explanatory variables (i.e., various defect features) by performing variable selection and regularization simultaneously; decision trees generate rule-based relationships in a tree structure by splitting the dataset on decision nodes according to the values of the explanatory variables; random forests construct a multitude of decision trees and output a mean regression value of the individual trees.

The experiment is based on 5-fold cross-validation to train these machine learning models using datasets 1, 2, and 3, thereby avoiding overfitting and improving the generalizability of the models shown in Table 3.7.

Table 3.7: Comparisons of modeling error of different ML methods via 5-fold cross-validation in predicting fatigue life from defect features in terms of MAPE for testing data for three datasets. Note: the MAPE in the table is composed of average MAPE of 5-fold datasets with its standard deviation.

Dataset	LASSO	Decision trees	Random forest	SVR
Dataset 1	0.891 ± 0.0461	0.948 ± 0.1292	0.800 ± 0.0924	0.633 ± 0.0477
Dataset 2	0.541 ± 0.1802	0.454 ± 0.1192	0.385 ± 0.1201	0.336 ± 0.1049
Dataset 3	0.651 ± 0.1942	0.575 ± 0.3515	0.578 ± 0.2212	0.573 ± 0.2037

It is noted in Table 3.7 that SVR outperforms the other three benchmark methods with the smallest MAPE and less variance for all three datasets. Decision trees have the largest MAPE values for dataset 1, indicating the rule-based method cannot accurately represent the relationships between the defect features and the fatigue life in different heat treatments and applied strain amplitudes. LASSO also underperforms since the relationships between the defect features and fatigue life are unlikely to be linear. As an ensemble learning method of decision trees, random forests gain modeling accuracy improvement by approximating sophisticated nonlinear relationships, and have relatively small MAPE values, especially for Dataset 2. It indicates that

SVR is capable of well predicting the fatigue life of L-PBF specimens.

From the data perspective, Dataset 2 and Dataset 3 are generated under consistent conditions (i.e., Dataset 2 includes L-PBF specimens tested under the same strain amplitude, i.e., 0.003 mm/mm, Dataset 3 includes L-PBF specimens that underwent the same heat treatment and tested under the same strain amplitude); therefore, despite the small data sizes, the possible clear patterns in the datasets under the consistent conditions enable accurate modeling of the relationships between defect features and fatigue life. On the contrary, due to the different heat treatment and strain values applied, Dataset 1, with all the specimens, has inferior modeling accuracy and is not favorable for analyzing defect criticality despite its relatively large data size.

Table 3.8: SVR modeling error under different hyperparameters of all datasets via 5-fold cross-validation. Note: γ controls the curvature of the decision boundary in the radial basis kernel; C controls the error in the objective function Eq. (5); Bold numbers represent the best MAPE.

Dataset	**Hyperparameters of SVR**		**Error Measurement**
	$\gamma (> 0)$	C	MAPE
Dataset 1	0.5	4	0.224
Dataset 2	0.5	4	**0.101**
Dataset 3	0.5	4	0.120

Then this work selects the best SVR hyperparameters from a 5-fold cross-validation grid search to train all datasets and obtain the best prediction error shown in Table 3.8. These SVR results are used to interpret the impact of defect features on the predicted fatigue life by model-agnostic methods in the next section.

Model interpretability and result discussion

In the IDADC, this work leverages the model-agnostic methods (PFI and ALE) to enhance the result interpretations for SVR. Specifically, this work firstly adopts PFI to evaluate the marginal importance of defect features towards the SVR outcomes of fatigue life. Then, this work uses ALE to further examine how important defect features influence the expected outcomes of SVR. Such methods can enhance the interpretability of SVR and increase the understanding of the relationships between critical defect features and fatigue life of L-PBF specimens from ML.

PFI measures the importance of a feature by calculating the increase in the model error after permuting the feature. They are illustrated in Figure 3.9 for the three datasets in this study. As shown in Figure 3.9, strain amplitude and normalized stress level parameter (adjusted for heat treatment) are dominant features in Dataset 1 since Dataset 1 includes all specimens with different heat treatments and applied strain amplitudes. The results align with the previous correlation analysis in Section 5.2.1. Again, it indicates that due to the inconsistent heat treatment and strain amplitudes, Dataset 1 is not favorable for defect criticality analysis despite its relatively large data size. For Dataset 2 and Dataset 3, while large variations exist in the importance of the defect features, they are not statistically significant. ALE is used to interpret the importance of defect features in Dataset 2 and Dataset 3.

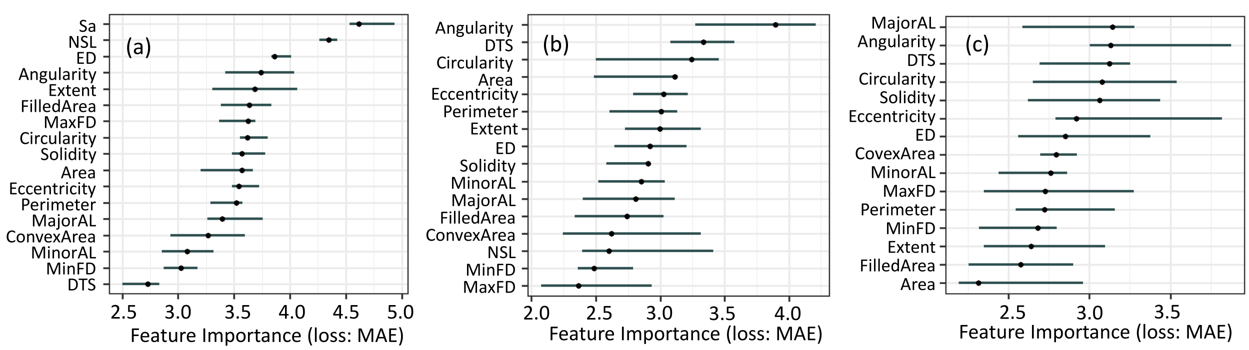

Figure 3.9: PFI of defect features based on SVR for L-PBF 17-4 PH SS specimens in (a) Dataset 1, (b) Dataset 2, and (c) Dataset 3. (NSL: Normalized stress level; DTS: The distance to the surface; ED: Equivalent diameter; MaxFD: Max Feret diameter; MinFD: Min Feret diameter; MajorAL: Major axis length; MinorAL: Minor axis length)

ALE calculates the marginal effects of the interested features on the fatigue life modeled with SVR by marginalizing other features from its distribution function as in Eq. (6) and Eq. (7) and describes the effects by using the incremental/decremental differences to the fatigue life, indicating as upward/downward trend in ALE plots. ALE plots are used to analyze some important size-related, morphology-related, and location-related defect features identified from PCC, MIC, and PFI above, for Dataset 2 and Dataset 3. Moreover, this work further investigates the defects on fractography for explanations of abnormal sections in the general trend on the ALE plots by considering the synergistic effect of the various features on fatigue lives.

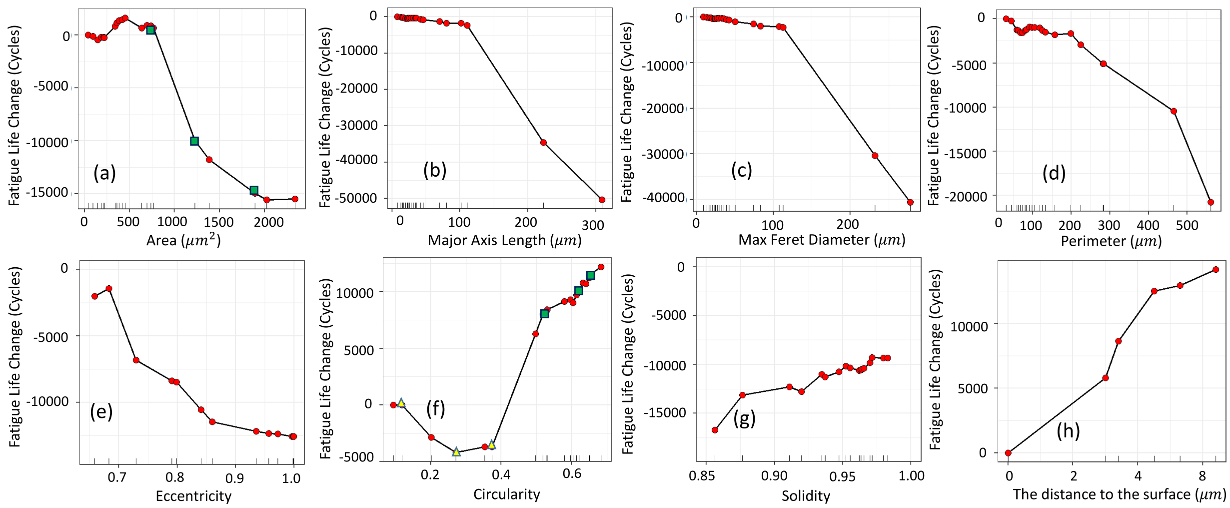

Figure 3.10: The ALE plots of defect features for SVR, exhibiting their impacts on fatigue life of L-PBF 17-4 PH SS specimens for Dataset 2 (Yellow triangle: abnormal points not on trend; Green rectangle: sampling points on the right trend).

(1) ALE plot interpretability and discussion for Dataset 2. Generally speaking, ALE plots for the eight defect features in Dataset 2 show clear trends regarding fatigue lives due to its relatively large data size, as in Figure 3.10. It is noted that the increase of all the size-related features (e.g., Area, Major Axis Length, Max Feret Diameter, and perimeter) lead to the decrease of fatigue lives of L-PBF specimens as shown in Figure 3.10 (a)-(d). For instance, an examination of defects for Area in Figure 3.10 (a) shows that the defect area only becomes critical after reaching a certain threshold value (around 700–800 μm^2), after which the fatigue lives significantly deteriorate with an increase in size. The defects above the threshold can be classified as large, irregularly shaped LOFs shown in Figure 3.11 below, while the defects below the threshold could be small gas-entrapped pores. The cracks that grow out of defects larger than 700–800 μm^2 may have already become a long crack, and its growth rate may have been high, which results in shortening the total crack propagation life. Therefore, there is a stronger correlation between life and larger defect sizes. Additionally, in Figure 3.10 (b)-(d), the general trends show that the fatigue lives decrease with an increase in the values of other size-related features (i.e., major axis length, max Feret diameter, and perimeter).

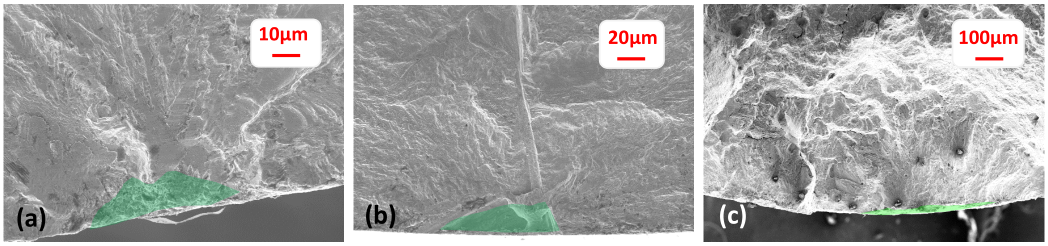

Figure 3.11: Fatigue crack initiating defects correspond to the three defects with a large area (>700–800 μm^2) and having a large decrease in fatigue lives in Figure 6 (a) marked in green rectangles. They show large size and irregular shapes and could be classified as LOFs.

Furthermore, the general trends of morphology-related defect features (e.g., eccentricity, circularity, solidity) also align with fatigue mechanisms and influencing factors and are verified by PCC in Table 5 as well. For instance, the fatigue lives of specimens tend to be longer if the critical defects are more circular (circularity), have a larger radius of curvature (eccentricity), or/and is more convex with a smoother contour (solidity) if size-related features and distance to the surface are the same. It is noted in Figure 3.10 (f) that while most defects with circularity (> 0.5) show an upward trend that is positively correlated with fatigue lives, several others with small circularity (< 0.5) (marked as yellow triangles) have a negative correlation with fatigue lives. This work evaluated these defects and found they are irregular (LOFs) and are open to the surface, as shown in Figure 3.12 (a)-(c) below. Except for these LOFs, it can be seen from Figure 3.12 (d)-(f), the fatigue lives show an upward trend with the increase of circularity.

Lastly, it is worth noting that the fatigue lives of LB-PBF specimens in Dataset 2 are longer if the critical defects are far away from the surface (i.e., having a large distance to the surface), as shown in Figure 3.10 (h).

(2) ALE plot interpretability and discussion for Dataset 3. As for Dataset 3, firstly, it is observed that the increase of some size-related features (such as area, major axis length, max Feret diameter) will generally decrease the fatigue lives of L-PBF specimens, as shown a patent downward trend in the above Figure 3.13 (a)-(c). Such observations align with known fatigue mechanisms and influencing factors and have been reported in Table 5 with negative correlations by PCC. Interestingly, as seen for some of the size-related features, the defect area in Figure 3.13

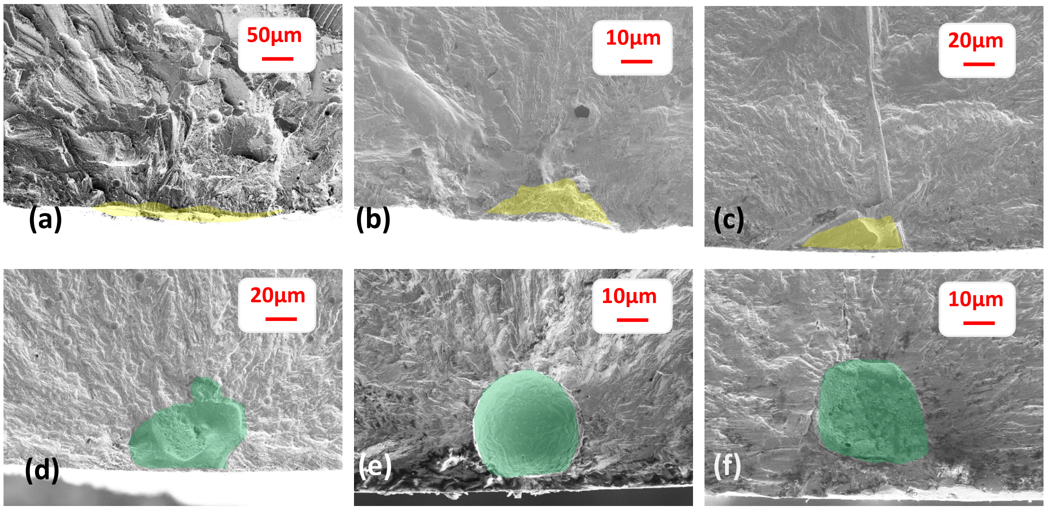

Figure 3.12: Fatigue crack initiating defects from Dataset 2 for L-PBF 17-4 PH SS specimens with varying circularity values and fatigue lives. As seen in the Fatigue Life vs. circularity ALE plot in Figure 6 (f), the defects with circularity (< 0.5) and marked as yellow triangles are defects in (a)-(c), and the defects with circularity (> 0.5) and marked as green rectangles are defects in (d)-(f).

(a) only becomes critical after reaching a certain threshold value (around 700–800 μm^2), after which the fatigue lives deteriorate with an increase in size. Inspecting these defects revealed that below this threshold area value, the gas-entrapped pores, which typically are small and circular, are responsible for crack initiation; the defects above the threshold can be classified as large, irregularly shaped LOFs shown in Figure 10 below. The cracks that grow out of defects larger than 700–800 μm^2 may have already become a long crack; its growth rate may have been high, shortening the total crack propagation life. This observation is very similar to Dataset 2 in the previous section. Additionally, in Figure 3.13 (b) and (c), the general trends show that the fatigue lives decrease with an increase in the values of other size-related features (i.e., major axis length, max Feret diameter).

The relationships between perimeter (also a size-related feature) and fatigue live in Figure 3.13 (d) do not show a clear trend. As shown in Figure 3.13 (d), while the defects with a larger perimeter ($> 100\mu m$) are on the downward trend of fatigue lives with the increase of perimeter, the defects with a smaller perimeter ($< 100\mu m$) have an unusual dip in fatigue lives (a huge decrease and then a rebound in fatigue lives). Three defects at this dip (marked as yellow triangles

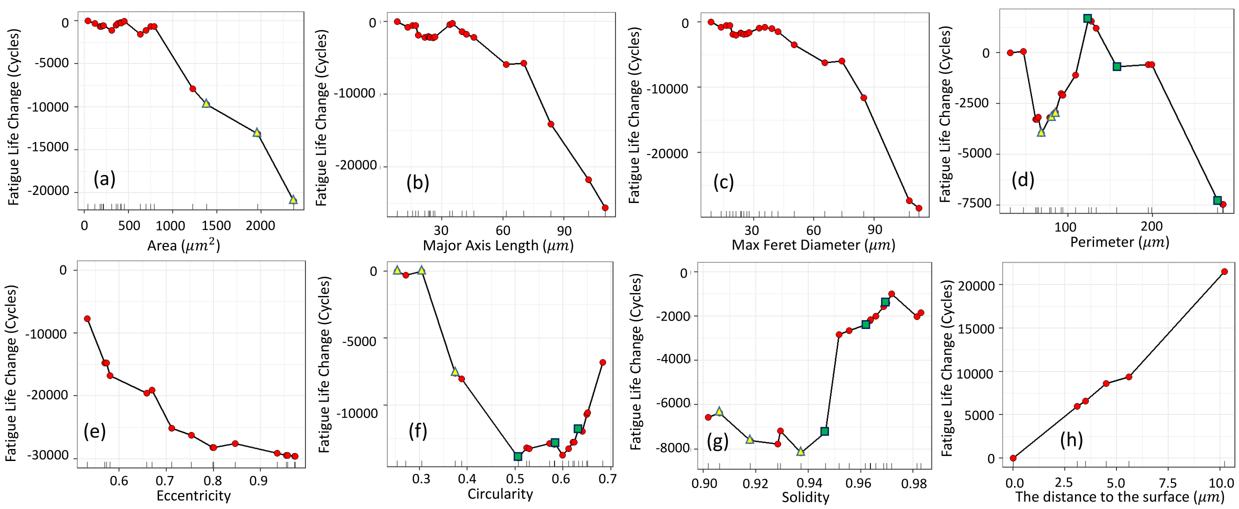

Figure 3.13: The ALE plots of defect features for SVR, exhibiting their impacts on fatigue life of L-PBF 17-4 PH SS specimens for Dataset 3 (Yellow triangle: abnormal points not on trend; Green rectangle: sampling points on the right trend).

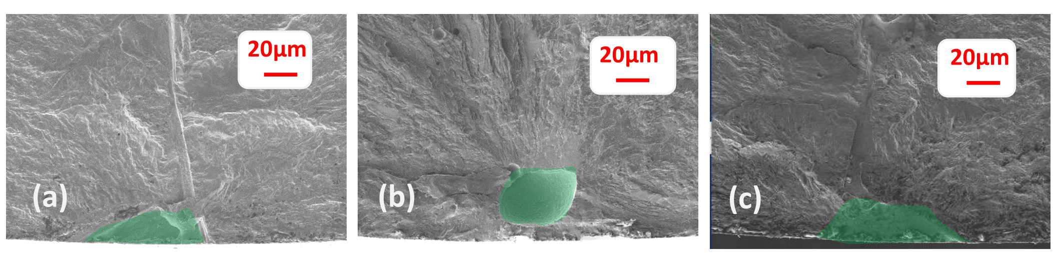

Figure 3.14: Fatigue crack initiating defects correspond to the three defects with a large area (>700–$800\mu m^2$) and having a large decrease in fatigue lives in Figure 3.13 (a), marked in green rectangles. They show large size and irregular shapes and could be classified as LOFs.

in Figure 3.13 (d)) are identified and shown in Figure 3.15 (a)-(c) below. The sharp drop in the fatigue lives with small perimeters could be attributed to their locations (open to the surface), as shown in Figure 3.15 (a) and (b). When the defects with a small perimeter move away from the surface, as shown in Figure 3.15 (c), the fatigue lives rebound. Moreover, three defects after the dip on the downward trend are sampled, marked as green rectangles in Figure 3.13 (d), and shown in Figure 3.15 (d)-(f) here. Their increasing perimeters can correlate with the decreasing trend of fatigue lives. Additionally, as shown in Figure 3.15 (f), the defect with the largest perimeter ($\sim 300\mu m$) and open to the surface has a substantial decrease in the fatigue life.

Secondly, it is noticed in Figure 3.13 (e) that as the eccentricity value increases, the fatigue lives decrease, since a large eccentricity means the defect has a long major axis length and it is noncircular. This feature is strongly correlated with major axis length and the maximum Feret diameter, which is also reflected in the similar trends observed for these size-related features. Similarly, for defects with high values of circularity and solidity, fatigue lives are higher. This is due to the defects having smoother contours, a larger radius of curvature, and a smaller size, often, which is accompanied by a reduction in the local stress concentrations, and thus, leading to better fatigue lives. This work further examines the impact of circularity and solidity on fatigue life in the following two paragraphs since there are some abnormal sections in the general trend on the ALE plots with Dataset 3 in Figure 3.13 (f)-(g).

PCC in Table 5 suggests that the fatigue lives of specimens with large circularity values tend to be long. Defects with high circularity (> 0.5) in Figure 3.13 (f) do show an upward trend with positive correlations with fatigue lives. However, a few defects with small circularity (< 0.5) (marked as yellow triangles in Figure 3.13 (f)) indicate abnormal negative correlations. This work identifies these defects and presents them in Figure 3.16 (a)-(c) below. These defects are irregular (LOFs) and are open to the surface, which will have largely decreased fatigue lives; especially for the defect in Figure 3.16 (c), it has a large area of 1382.667 μm^2. After the dip, it can be seen from Figure 3.16 (d)-(f), the fatigue lives are on the upward trend with the increase of circularity.

PCC in Table 5 also suggests the fatigue lives of specimens with large solidity values tend

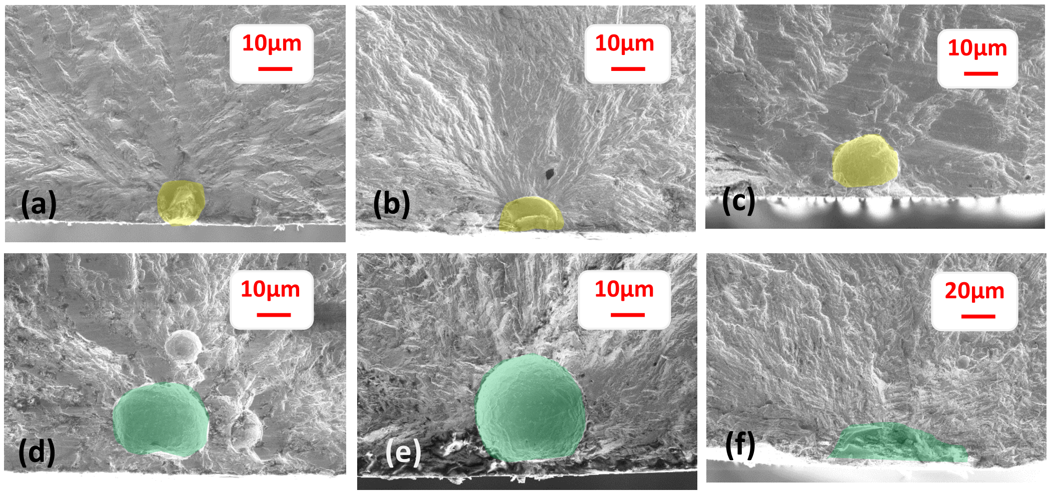

Figure 3.15: Fatigue crack initiating defects from Dataset 3 for L-PBF 17-4 PH SS specimens with varying perimeter values and fatigue lives. As seen in the Fatigue Life vs. Perimeter ALE plot in Figure 3.13 (d), the defects with a small perimeter ($< 100\mu m$) and marked as yellow triangles are defects in (a)-(c), and the defects with a large perimeter ($> 100\mu m$) and marked as green rectangles are defects in (d)-(f).

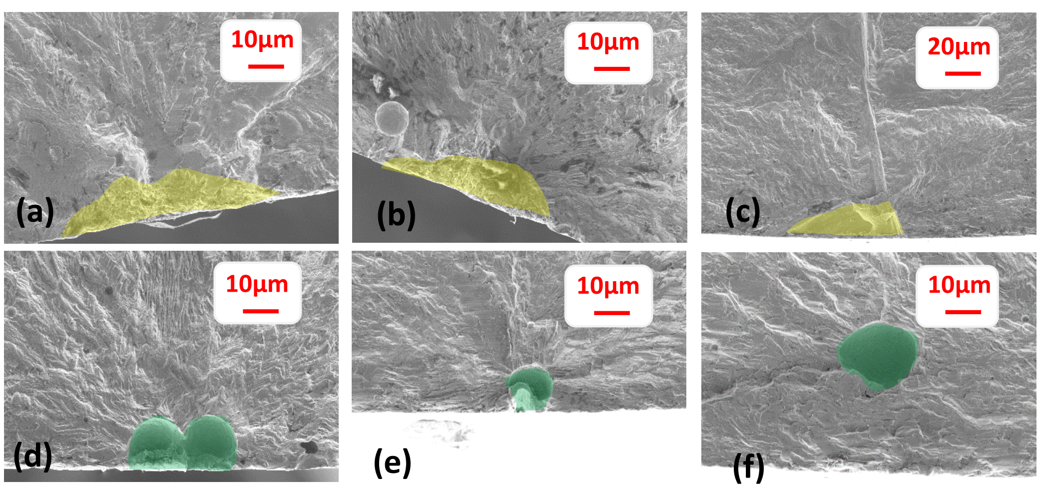

Figure 3.16: Fatigue crack initiating defects from Dataset 3 for L-PBF 17-4 PH SS specimens with varying circularity values and fatigue lives. As seen in the Fatigue Life vs. circularity ALE plot in Figure 3.13 (f), the defects with circularity (< 0.5) and marked as yellow triangles are defects in (a)-(c), and the defects with circularity (> 0.5) and marked as green rectangles are defects in (d)-(f).

to be long because a large solidity indicates the defect is close to a convex shape and might have a smooth boundary and be gas-entrapped pores. This aligns with the upward trend in Figure 3.13 (g) except for a few defects with solidity less than 0.94 having an excessive reduction in fatigue lives. These defects (marked as yellow triangles in Figure 3.13 (g)) are shown in Figure 3.17 (a)-(c), and it is noticed that they are open to the surface and their areas are large, relative to defects (marked as green rectangles in Figure 3.13 (g)) in Figure 3.17 (d)-(f).

Finally, Figure 3.13 (h) shows that defects with a shorter distance to the surface lead to shorter fatigue lives of L-PBF specimens. This can be attributed to the presence of higher stress concentrations when defects are close to the surface, as has also been seen through FE modeling in literature, where the highest stress concentrations were found for defects less than one diameter away from the free surface [58].

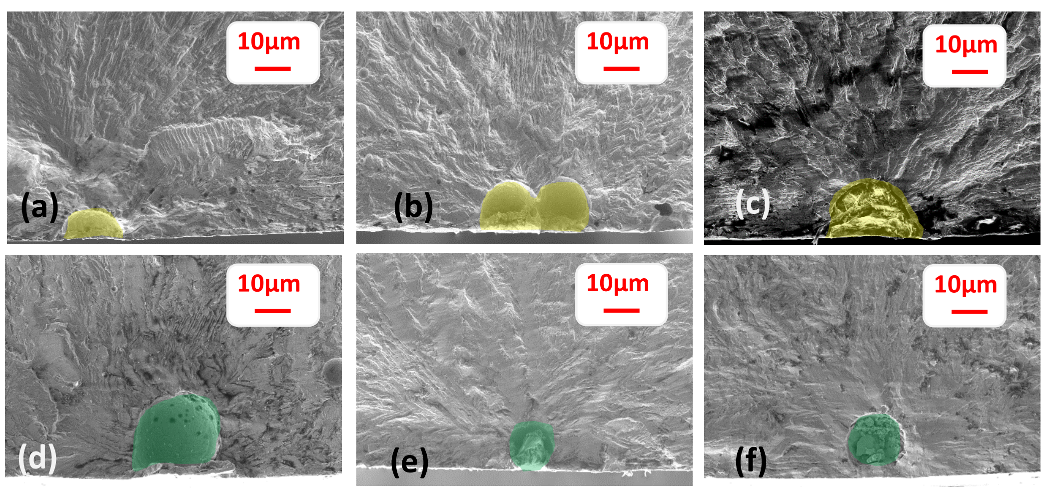

Figure 3.17: Fatigue crack initiating defects from Dataset 3 for L-PBF 17-4 PH SS specimens with varying solidity values and fatigue lives. As seen in the Fatigue Life vs. solidity ALE plot in Figure 3.13 (g), the defects with solidity (< 0.94) and marked as yellow triangles are defects in (a)-(c), and the defects with circularity (> 0.94) and marked as green rectangles are defects in (d)-(f).

In summary, ALE plots provide a promising analytical way to explore and interpret the correlations between defect features and fatigue lives of L-PBF specimens. In practice, especially when data are limited, combining both size and morphology-related features and considering the synergistic effect of the various features on fatigue lives can give better interpretations.

3.5 Results of Case Study 2: Fatigue Life Prediction and 3D Critical Directional Curvature Feature Analysis of L-PBF Parts

Based on the destructive defect criticality analysis and feature correlation analysis, the proposed critical directional curvature of 3D XCT critical defects was evaluated in three scenarios for the nondestructive prediction of fatigue life in L-PBF parts. Table 3.9 and Figure 3.19 show the performance of stress-adjusted fatigue life prediction of L-PBF parts using multiple regression models.

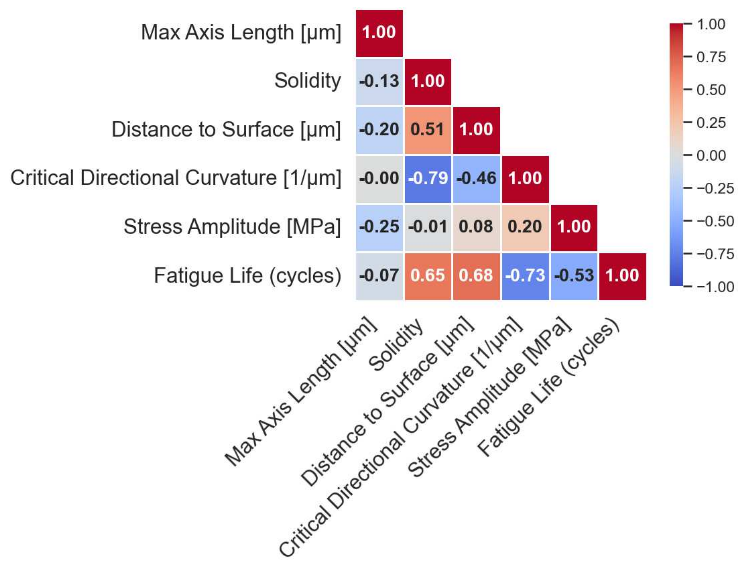

Figure 3.18: Spearman correlation matrix of input features utilized for nondestructive fatigue life prediction based on 3D defect characteristics. The critical directional curvature shows a strong correlation (-0.73) with fatigue life, indicating it captures fatigue-relevant defect geometry. Moderate correlations are observed among some geometric descriptors, while no significant multicollinearity is shown.

As shown in Table 3.9, for both linear and nonlinear models, incorporating the proposed critical directional curvature feature in the curvature-enhanced case significantly enhances prediction accuracy compared with the baseline case, which only includes critical defect size and distance to the surface. Importantly, the simple linear regression model already achieves a low RMSE of 0.2770 (a 22.46% improvement), a high adjusted R^2 of 0.7367 (a 22.38% improvement), and a low SMAPE of 0.5804 (a 16.48% improvement). These improvements indicate the critical

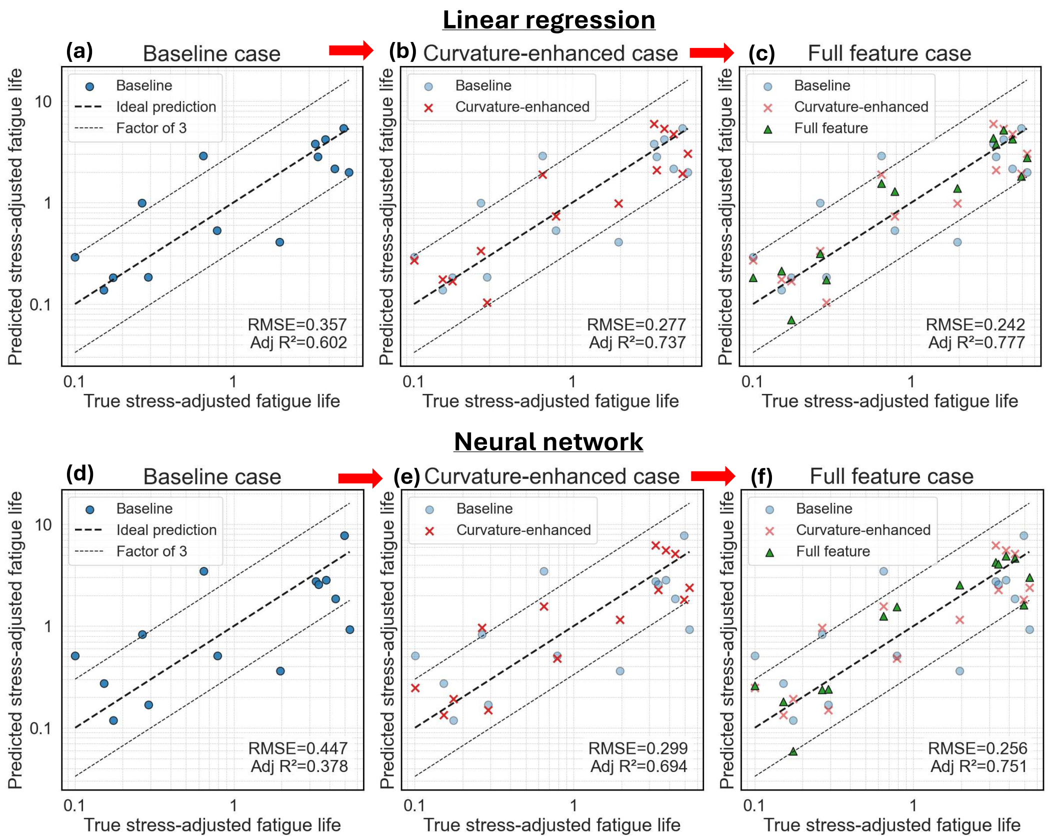

Figure 3.19: Comparison between predicted and actual stress-adjusted fatigue life for linear regression and NeuralNet models. (a)–(c) present the results of the linear regression model, identified as the best-performing nondestructive prediction model, showing a significant improvement after incorporating critical directional curvature in case (b) compared with the baseline case in (a), with an adjusted R^2 of 0.737. (d)–(f) illustrate the prediction results of the neural network model, where adding directional curvature in case (e) enhances prediction accuracy relative to the baseline case in (d), yielding an adjusted R^2 of 0.694.

directional curvature captures the local stress concentration of critical defects, consistent with notch theory ($K_t \propto 1/\sqrt{r} \propto \sqrt{k}$), where higher curvature causes increased stress and reduced fatigue life. By considering both the defect geometric sharpness and the loading direction, critical directional curvature offers a physically interpretable feature that influences crack initiation behavior. Interestingly, the linear regression model outperformed NeuralNet in nondestructive fatigue life prediction despite its simpler structure and achieved a similar performance as Gaussian process regression. It indicates that the relationship between critical directional curvature and fatigue life is mainly linear, with the removal of the stress amplitude effect via

Table 3.9: Comparison of regression models for nondestructive fatigue life prediction among three cases. The bold values highlight the significant prediction accuracy improvement of the curvature-enhanced case, compared to the baseline case.

Model	Scenario	RMSE	Gain (%)	$Adj\,R^2$	Gain (%)	SMAPE	Gain (%)
Linear regression	1. Baseline case	0.3572	–	0.6020	–	0.6949	–
	2. Curvature-enhanced case	**0.2770**	**22.46↓**	**0.7367**	**22.38↑**	**0.5804**	**16.48↓**
	3. Full feature case	0.2419	32.26↓	0.7768	29.04↑	0.5716	17.73↓
Ridge regression	1. Baseline case	0.3604	–	0.5947	–	0.7837	–
	2. Curvature-enhanced case	**0.2714**	**24.69↓**	**0.7472**	**25.63↑**	**0.5288**	**32.53↓**
	3. Full feature case	0.2434	32.48↓	0.7742	30.17↑	0.5236	33.18↓
NeuralNet	1. Baseline case	0.4465	–	0.3779	–	0.8431	–
	2. Curvature-enhanced case	**0.2987**	**35.34↓**	**0.6938**	**83.57↑**	**0.7104**	**18.58↓**
	3. Full feature case	0.2558	44.63↓	0.7505	98.59↑	0.5478	37.23↓
Gaussian process regression	1. Baseline case	0.3590	–	0.5979	–	0.6944	–
	2. Curvature-enhanced case	**0.2743**	**23.59↓**	**0.7418**	**24.06↑**	**0.5596**	**19.42↓**
	3. Full feature case	0.2681	25.32↓	0.7259	21.40↑	0.6017	13.36↓

the physics-informed modeling. This finding supports that the developed approach achieves accurate nondestructive fatigue life prediction through physically meaningful features. While additional defect morphology features, such as solidity in the full feature case, showed minor gains, the major improvement arises from adding critical directional curvature, which effectively links critical defect geometric sharpness and fatigue life of L-PBF parts.

These results are further supported by the feature correlation analysis using Spearman correlation, shown in Figure 3.18. The critical directional curvature shows a strong correlation of -0.73 with fatigue life, while no severe multicollinearity is observed among the input features. This indicates that the critical directional curvature captures fatigue-relevant defect geometry, which serves as a dominant and physically meaningful feature for fatigue life.

As shown in Figure 3.19, both regression models demonstrate a clear improvement in stress-adjusted fatigue life prediction accuracy after incorporating the critical directional curvature ((b) and (e)) compared with the baseline case ((a) and (d)). The scatter of the prediction is significantly reduced. Moreover, a minor improvement is observed in the full feature case ((c) and (f)), indicating that most of the predictive improvement is from the curvature feature. These results are consistent with the quantitative findings in Table 3.9 and confirm that the addition of the critical directional curvature feature enables accurate nondestructive fatigue life prediction and physically interpretable results.

Table 3.10: Linear regression coefficients and significance of defect features for nondestructive stress-adjusted fatigue life prediction. It demonstrates that the defined critical directional curvature exhibits a negative and statistically significant coefficient to stress-adjusted fatigue life.

Feature	**Coefficient**	t**-value**	p**-value**	**Significance**
Distance to surface (μm)	$+0.88$	$+4.73$	0.001	Highly significant
Solidity	$+0.54$	$+2.91$	0.020	Significant
Critical directional curvature ($1/\mu m$)	**-0.48**	**-3.20**	**0.013**	**Significant**
Max axis length (μm)	-0.43	-2.24	0.055	Marginally significant

To further explore the feature importance and their relationships with stress-adjusted fatigue life, this work investigated linear regression coefficients and employed the NeuralNet SHAP values in the full feature case. As shown in the Table 3.10, the defined critical directional curvature exhibits a negative and statistically significant coefficient (-0.48, $t = -3.2$, $p = 0.013$), indicating that sharper local geometry in the critical defect with higher curvature leads to reduced fatigue life. This finding aligns with fracture mechanics, which suggests that sharp regions increase stress concentration and promote early crack initiation. In comparison, conventional features, such as distance to surface and solidity, exhibit a relatively strong correlation with fatigue life; however, they do not capture the directional stress-related effect in the local geometry.

SHAP results in Figure 3.20 showed that the critical directional curvature consistently ranks as the third most important feature, comparable to solidity. It highlights the significance of critical directional curvature as a key feature in fatigue prediction. It shows a consistent negative relationship between critical directional curvature and stress-adjusted fatigue life (higher

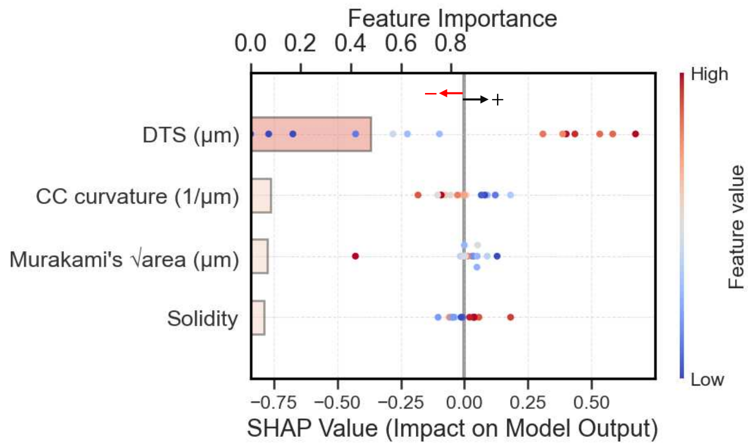

Figure 3.20: This SHAP result shows that the critical directional curvature was identified as the third most influential feature from the left feature importance comparison. In addition, it reveals that critical directional curvature has a strong negative correlation with the predicted stress-adjusted fatigue life, which is consistent with the results in Table 3.10. Note: CD curvature is the critical directional (CD) curvature.

curvature, lower SHAP values, reduced predicted fatigue life), indicating that critical defects in an L-PBF specimen with higher directional curvature (sharper geometry) lead to more stress concentration and reduce fatigue life. The relationships of other features, such as maximum axis length, distance to the surface, and solidity, are also shown to have consistent physical relationships with fatigue life, where larger and irregular critical defects that are close to the surface of the specimen reduce fatigue life. In the linear regression analysis (Figure 3.19b), incorporating the curvature feature (curvature-enhanced case and full feature case) reduces the coefficient magnitude of distance to the surface and solidity, which indicates that the curvature feature explains the partial variance attributed to these features, as curvature inherently encodes local geometry and orientation information of critical defects.

To further validate the physics interpretation of the proposed critical directional curvature, Figure 3.21 shows the highest directional curvature and the estimated highest stress concentration factor of the critical defect in spatial correspondence for these three representative specimens. The directional curvature map (middle row) highlights the convex region ($k_1 > 0, k_2 > 0$) with maximum directional curvatures aligned with the loading direction, while the stress concentration factor map (bottom row) demonstrates the same location with maximum

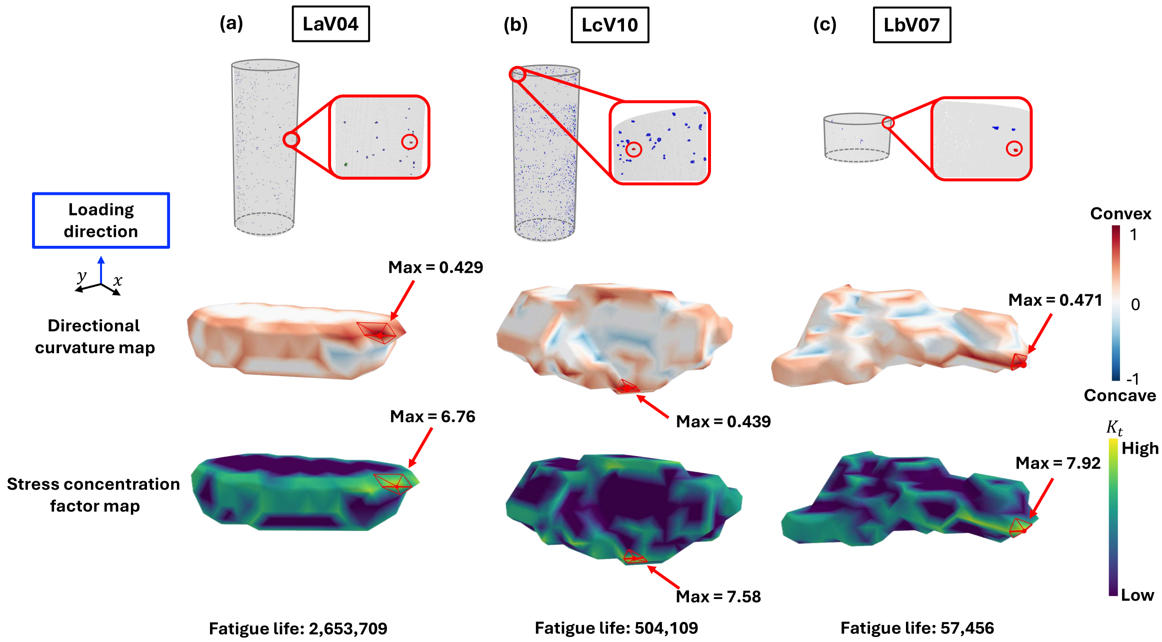

Figure 3.21: Three corresponding 3D XCT critical defects with specimens illustrate the local geometry with high directional curvature (middle row), indicating that regions with higher positive curvature exhibit sharper convex geometry. The stress concentration factor maps (bottom row) demonstrate that areas with higher local directional curvature show increased stress concentration.

SCF as the directional curvature map. It confirms that the proposal curvature feature identifies the fatigue-critical local regions estimated by notch theory. Moreover, this work observed a clear trend among these three specimens with LoF critical defect that with the increase of maximum directional curvature (0.429, 0.439, and 0.471) with corresponding K_t (6.76, 7.58, and 7.92), the experimental fatigue life decreases (2.65×10^6, 5.04×10^5, and 5.75×10^4 cycles). This relationship indicates that sharper defect geometries result in higher local stress concentrations and earlier crack initiation.

Based on the results, critical direction curvature indicates that sharper defect geometries with high curvature increase local stress concentration and accelerate the initiation of cracks, as validated by critical contour curvature in destructive fractography analysis. They all exhibit a negative correlation with stress-adjusted fatigue life, as confirmed by both data-driven and simulation validation, indicating that the proposed critical directional curvature serves as a physics-based descriptor of defect severity for nondestructive fatigue life prediction in L-PBF.

3.6 Summary

In this study, an IDADC for L-PBF fatigue performance assessment from critical defect features and a new 3D critical directional curvature feature were developed. IDADC is a novel integration of domain expertise and data-driven methods for potential qualification of L-PBF components. The proposed IDADC incorporates defect feature extraction from fractography images of L-PBF fatigue test specimens, correlation analysis, and defect-fatigue relationships modeling with ML to achieve an accurate and interpretable data-driven fatigue analysis. Specifically, based on SEM images acquired from the fracture surfaces of L-PBF 17-4 PH SS specimens tested under strain-controlled, fully reversed fatigue loading, CBIR is applied to extract various size-related and morphology-related defect features from the critical defects. Then, correlation analysis and SVR are employed to understand the relationships between these defect features and the fatigue life of the specimens. The analysis reveals the correlations between critical defect features and the fatigue life, which align with the known fatigue mechanisms and influencing factors. To summarize the findings of case study 1:

- Defect size, distance to the surface, circularity, and smooth contours were found to have strong relationships with the fatigue life of L-PBF specimens. SVR achieved relatively accurate fatigue life modeling based on defect features, especially with the specimens that had undergone the same heat treatment and were tested under the same strain amplitude (MAPE = 0.101).

- PFI and ALE were incorporated to interpret the relationships from SVR outcomes, revealing morphology-related features (such as eccentricity, circularity, and solidity) and size-related features (such as maximum ferret diameter, perimeter, and area) have a significant impact on the fatigue life of L-PBF specimens. They provided useful, interpretable insights from "black-box" ML models.

And the main results and findings through the designed case study 2 for stress-adjusted fatigue life (i.e., removing the effect of stress amplitude) prediction in L-PBF Ti-6Al-4V parts were as follows:

- Incorporating the critical directional curvature feature significantly improved nondestructive stress-adjusted fatigue life prediction by 22.46% (RMSE) in linear regression and 35.34% in NeuralNet prediction, compared with models using only conventional defect size and location features.
- Beyond prediction, critical directional curvature enabled the localization of potential high stress concentration regions in nondestructive XCT. They provide physics validation of fatigue-critical sites and physical interpretation for fatigue life prediction.
- SHAP and simulation analysis in fatigue life prediction revealed a consistent negative correlation between curvature features and fatigue life, which aligns with the physical understanding that sharper local geometry (higher curvature) of a critical defect induces higher local stress, accelerates crack initiation, and reduces fatigue life.

Overall, the IDADC has the potential practical benefits to nondestructive inspections with 3D X-ray CT scans in the future, which will become the mainstream in describing defects in L-PBF specimens [12, 14–16, 169] and studying L-PBF fatigue properties:

1. To identify the critical defects in L-PBF specimens from 3D X-ray CT scans: X-ray CT scans can capture the size-related, morphology-related, and distance-related defect features for all the defects in the scanned areas. The IDADC and the insights from this study can be easily extended to the nondestructive defect inspection and identify the critical defects that may initiate cracks in the parts during cyclic service loading. Such an understanding will potentially enable a nondestructive qualification of additively manufactured parts for fatigue-critical applications by utilizing an X-ray CT and scanning the critical locations of the parts that experience higher stress.
2. To understand the most critical defect features of the different types of defects (i.e., LOFs and keyholes) in AM materials in 3D, correlate them to the L-PBF process parameters, and

study their impacts on the fatigue behavior of AM materials. The insights obtained can be used to optimize process parameters to avoid such defect types and/or reduce the intensity of such features, therefore improving the fatigue lives of the AM materials.

3. To reinforce the fidelity of defect features with strong correlations: Some features are more sensitive to X-ray CT inspection resolutions (e.g., circularity) than others (e.g., volume, major axis length). The strong correlations among these features can be used to adjust the measurements to a reasonable accuracy. For instance, circularity and major axis length have a strong correlation (~ 0.7), as shown in PCC analysis in Section 5.2.2. Since circularity is difficult to capture with low-resolution X-ray CT, this work can leverage its strong correlation with major axis length, which is still accurately measured with low inspection resolution, to adjust the measurement to a statistically more accurate circularity value. This benefit is particularly useful for 3D scanning for a large part with a limited resolution.

Chapter 4

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

Understanding the fatigue behavior and accurately predicting the fatigue life of laser powder bed fusion (L-PBF) parts remains a pressing challenge due to the complex nature of failure mechanisms, the time-consuming nature of tests, and the limited availability of 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.

4.1 Introduction

Understanding the fatigue performance of additively manufactured parts and assessing their fatigue life is 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 [64, 170]. 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 [171], 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 [172, 173]. Large volumetric defects primarily determine the mechanical properties [170], and they dominate the effect of microstructure anisotropy when their sizes are eight times the width of grains [174]. While post-process treatments can reduce the number of defects in L-PBF parts, they cannot fully eliminate them, particularly large or irregular defects [148, 175]. 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) [10], are generated under different energy inputs controlled by process parameters such as laser power and scanning speed, as shown in Figure 4.1 (a). 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 cracks 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 [12], as illustrated in Figure 4.1 (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 [10]. It is inspected from the fractography once the test specimen is broken, as shown in Figure 4.1 (c). Therefore, the defect features from XCT and fractography can be related to the tested fatigue life.

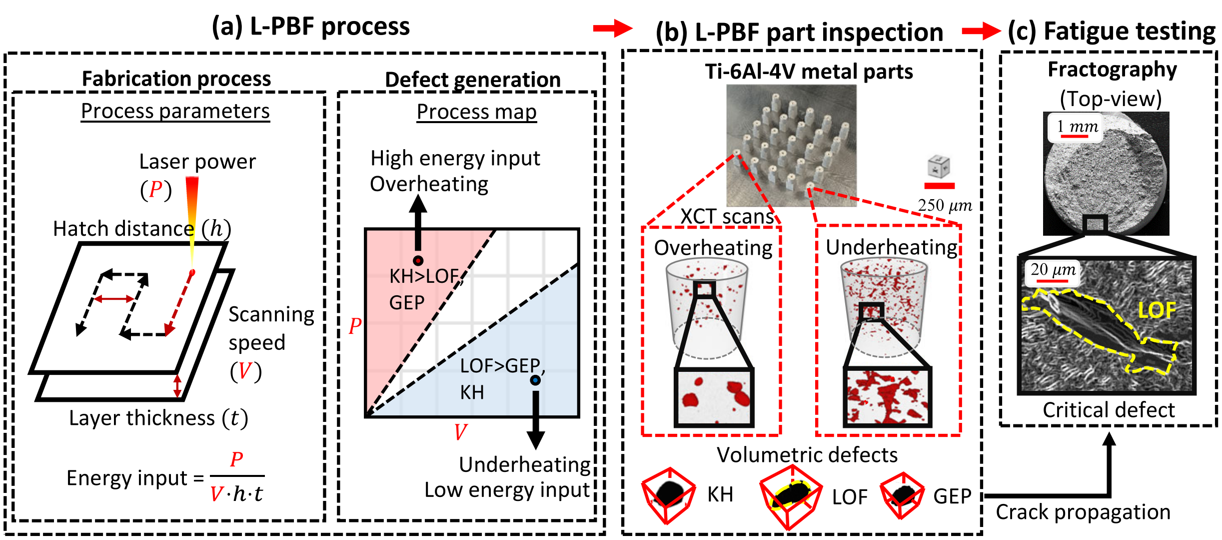

Figure 4.1: A process flow of L-PBF fatigue performance research from fabrication, process optimization, nondestructive inspection using XCT, and fatigue testing with fractography. (a) The energy input in fabrication is controlled by process parameters, leading to the co-occurrence of KHs, LOFs, and 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, a multimodal transfer learning (MMTL) framework is proposed 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 knowledge) 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 [176], 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.

4.1.1 Research gaps

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.

4.2 Research Experiments and Problem Formulation

4.2.1 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 4.1 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 4.1: 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

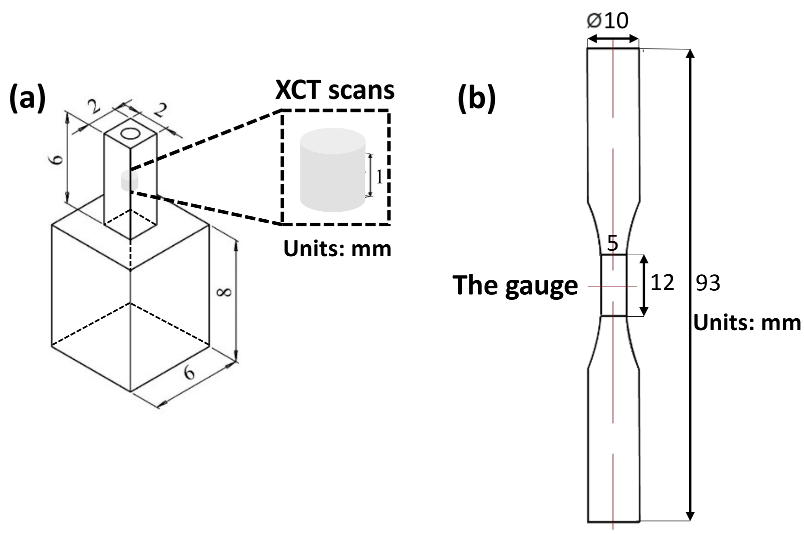

Figure 4.2: (a) The design geometry for the XCT scanning portion for Set 1, and (b) the XCT scanning portion at the gauge and during fatigue testing.

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 4.2 (a)) to permit high-resolution XCT scans (a ZEISS Xradia 620 Versa machine with 1 μm 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 [10]. 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 μm 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, a computer vision approach, content-based image retrieval [177], to extract diverse volumetric defect size- and morphology-related features by several shape descriptors was employed. Table 4.2 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 4.1) in volumetric defect classification has the potential to enhance the accuracy of classifications.

Table 4.2: 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 [178]), as shown in Figure

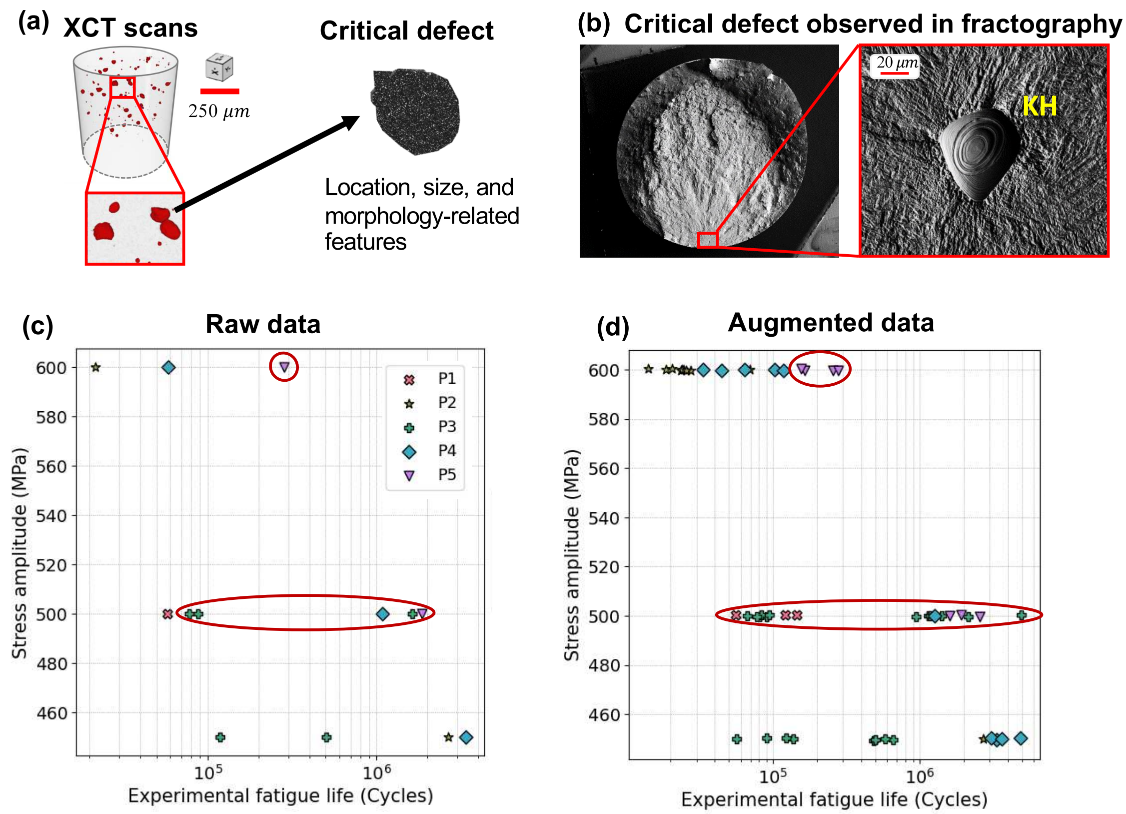

Figure 4.3: (a) The critical defect in XCT scans, (b) critical defect observed in fractography, (c) the raw stress-fatigue life (S-N) curve for L-PBF Ti-6Al-4V specimens (Set 2), fabricated under five distinct process conditions shown in Table 4.1, and (d) the augmented data (50 data points) from raw 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).

4.2 (b). This work 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 4.3 (b)) and then identified in the XCT scans (Figure 4.3 (a)) by a cross-dimensional defect matching process method [18] based on its location, size, and morphology. These features of the critical defects (e.g., length > 200 μm, aspect ratio < 0.3) are used to identify the potential critical defects in practical applications. [56] has shown this method can achieve 81.25% accuracy in correctly identifying critical defects from XCT scans. This work 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 4.3 (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 that some other factors have a larger impact on fatigue life, which are determined as critical defects from fractography in Figure 4.3 (b). Certain process conditions, such as P5, consistently demonstrate higher fatigue life across different stress levels.

Based on these insights and observations, the proposed MMTL framework designs a hierarchical structure incorporating defect features and process parameters into fatigue life prediction. To enhance the robustness of the results, this work 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, this work 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 4.3 (c) and (d).

4.2.2 Problem definition

Inspired by the insights gained from the aforementioned experiments, this work formulates 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. This work addresses this problem within the MMTL framework, as illustrated in Figure 4.4. 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 4.1). 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.

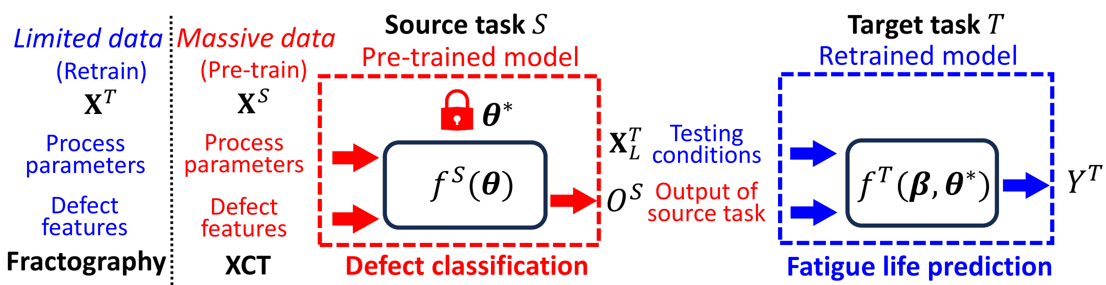

Figure 4.4: The formulation of the MMTL framework. A pre-trained model with massive XCT data for source task S defect classification and a retrained model with limited fractography data for target task T fatigue life prediction.

Problem formulation This work formulates the fatigue life prediction problem in the MMTL framework as

$$\begin{aligned} Y^T &= f^T(\boldsymbol{\beta}; \mathbf{X}_L^T, f^S(\boldsymbol{\theta}^*; \mathbf{X}^T)) \\ &= \underbrace{(f^T(\boldsymbol{\beta}; \mathbf{X}_L^T)}_{\text{Retrained term}} \circ \underbrace{f^S)(\boldsymbol{\theta}^*; \mathbf{X}^T)}_{\text{Pre-trained term}} \end{aligned} \tag{4.1}$$

where Y^T is the prediction of the target task T from the MMTL framework; $\mathbf{X}^T = [\mathbf{X}_1^T, ..., \mathbf{X}_M^T]^\intercal$ is the data with M modalities in the target task T (e.g., L-PBF process conditions and defects), and $\mathbf{X}_i^T = [X_{i,1}^T, ..., X_{i,d}^T]^\intercal$ 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); $f^T(\boldsymbol{\beta})$ is the TL function with the pre-trained model $f^S(\boldsymbol{\theta})$ with optimal parameters $\boldsymbol{\theta}^*$ 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 This work utilizes 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 $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ 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

$$\begin{aligned} &\min_{\boldsymbol{\beta}} \mathbb{E}[\mathcal{L}^T(Y^T, (f^T(\boldsymbol{\beta}; \mathbf{X}_L^T) \circ f^S)(\boldsymbol{\theta}^*; \mathbf{X}^T))] \\ &s.t.\ \boldsymbol{\theta}^* \in \arg\min_{\boldsymbol{\theta}} \mathbb{E}[\mathcal{L}^S(O^S, f^S(\boldsymbol{\theta}; \mathbf{X}^S))] \end{aligned} \tag{4.2}$$

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). $\mathbf{X}^S = [\mathbf{X}_1^S, ..., \mathbf{X}_M^S]^\top$ is the data for the source task S.

In this paper, this work considers 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.

4.3 Methodology of Multimodal Transfer Learning

4.3.1 Hierarchical graph convolutional network in source task

This work proposes 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 4.3.2). 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 4.5. GCN_1: $g_1(\theta_1)$ on Hierarchy 1 embeds process parameters to a latent space (i.e., defect feature space), while GCN_2: $g_2(\theta_2)$ on Hierarchy 2 focuses on learning defect features for defect classification.

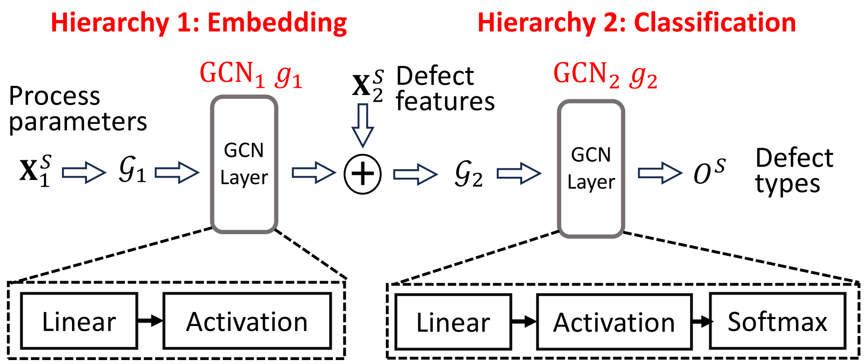

Figure 4.5: The proposed HGCN is designed for volumetric defect classification with two hierarchies. Hierarchy 1 encodes the first modality information in the graph-structured data, and Hierarchy 2 integrates the second modality and output from Hierarchy 1 to classify defects.

Graph representation of defects The data from all volumetric defects is organized into graphs, where nodes represent volumetric defects and edges represent node similarity ([179]). 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 $i = 1, ..., N$) 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 $j = 1, ..., N$) represents an individual defect j with the attribute of extracted defect features. An edge $e_{ii'} \in \mathcal{E}_2$ 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 [180, 181] 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. This work presents the comparable results from the simulation in Section 4.4, which utilized two overlapping components generated from a bivariate Gaussian mixture model. Besides, GCN leverages a 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 ($k = 1, ..., K$) can be represented as [182],

$$\mathbf{H}^{(k)} = \text{ReLU}(\tilde{\mathbf{A}}^{*}\mathbf{H}^{(k-1)}\theta^{(k-1)}) \tag{4.3}$$

where $\mathbf{H}^{(k)} = [h_1^{(k)}, ..., h_N^{(k)}]^\mathsf{T}$ 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 $\mathbf{D} = \text{diag}(d_1, ..., d_N)$ with $d_i = \sum_{v \in \mathcal{V}} \mathbf{A}^{*}[v_i, v_{i'}]$. $\theta^{(k)}$ is the weight matrix of the k-th layer.

Hierarchy 1: Embedding GCN_1 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),

$$g_1(\theta_1; \mathbf{X}_1^S) : \mathbf{H}_1^{(j)} = \text{ReLU}(\tilde{\mathbf{A}}_1^{*}\mathbf{H}_1^{(j-1)}\theta_1^{(j-1)}), \ j = 1, \dots, J \tag{4.4}$$

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 $\mathbf{H}_1^{(0)} = \mathbf{X}_1^S$, $\theta_1^{(j-1)} \in \mathbb{R}^{d \times q}$ is the weight matrix.

Hierarchy 2: Classification GCN_2 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

$$\begin{aligned} \mathbf{H}_2^{(k)} &= \text{ReLU}(\tilde{\mathbf{A}}_2^* \mathbf{H}_2^{(k-1)} \theta_2^{(k-1)}),\ k = 1, ..., K \\ g_2(\theta_2; \mathbf{X}_2^S) &: O^S = \text{Softmax} \circ \mathbf{H}_2^{(K)} \end{aligned} \tag{4.5}$$

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 $\theta_1^{(j-1)}$ follows $\mathbb{R}^{(m+q) \times p}$ since this work concatenates 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 θ_1 in Eq.(5.5) and θ_2 in Eq.(5.6), this work employs CE for the volumetric defect classification source task problem, as discussed in Section 4.2.2 Eq.(5.2),

$$\begin{aligned} \theta^* &\in \arg \min_{\theta = [\theta_1, \theta_2]} \mathbb{E}[-\sum_{c=1}^{C} O^S \log \hat{O}^S] \\ &= \arg \min_{\theta = [\theta_1, \theta_2]} \mathbb{E}[-\sum_{c=1}^{C} O^S \log (g_2(\theta_2; \mathbf{X}_2^S) \circ g_1)(\theta_2; \mathbf{X}_1^S))] \end{aligned} \tag{4.6}$$

This work opts for the Adam optimizer [183], a widely utilized optimization algorithm in deep learning, to update the parameters θ. The derivation of the parameter update for Eq.(5.7) can be found in Appendix 8.1.

4.3.2 Neural network layer in target task

Utilizing the pre-trained HGCN for the source task of defect classification, this work proceeds 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.(5.1) in Section 4.2.2),

$$Y^T = f^T(\boldsymbol{\beta}; \mathbf{X}_L^T, f^S(\theta^*; \mathbf{X}_1^T, \mathbf{X}_2^T)) = \sigma(\mathbf{Z}^{(l)}\boldsymbol{\beta}^{(l)}) \tag{4.7}$$

where σ is the activation function, $\mathbf{Z}^{(l)}$ is the output of l-th layer and the first layer input $\mathbf{Z}^{(0)} = [\mathbf{X}_L^T, f^S(\theta^*; \mathbf{X}_1^T, \mathbf{X}_2^T)] = [\mathbf{X}_L^T, \mathbf{H}_2^{(K),*}]$. $\boldsymbol{\beta}^{(l)}$ follows $\mathbb{R}^{(n+C)\times r}$ since this work concatenates fatigue-loading related features $\mathbf{X}_L^T$ (i.e., n dimensions) and output (last layer of GCN_2) of volumetric defect classification in source task $f^S(\theta^*; \mathbf{X}_1^T, \mathbf{X}_2^T)$ (i.e., C dimensions).

To determine the optimal $\boldsymbol{\beta}$ in Eq.(5.8), this work selects MSE for the fatigue life prediction target task problem as mentioned in Section 4.2.2 Eq.(5.2),

$$\begin{aligned} &\min_{\boldsymbol{\beta}} \mathbb{E}[\mathcal{L}^T(Y^T, (f^T(\boldsymbol{\beta}; \mathbf{X}_L^T) \circ f^S)(\theta^*; \mathbf{X}_1^T, \mathbf{X}_2^T))] \\ &= \min_{\boldsymbol{\beta}} \mathbb{E}[(\hat{Y}^T - Y^T)(\hat{Y}^T - Y^T)^\intercal] \end{aligned} \tag{4.8}$$

Here, this work also uses Adam optimizer [183] to update the parameters $\boldsymbol{\beta}$. The derivation can be found in Appendix 8.1.

4.4 Simulation Study

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

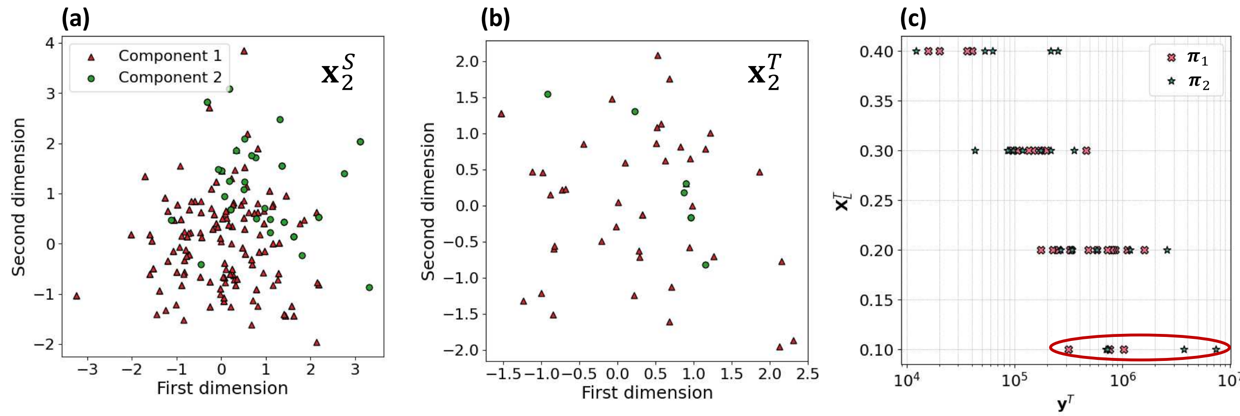

Figure 4.6: The visualization of source data (a) and target data (b) and simulated S-N curve (c). The simulated response $\mathbf{y}^T$ in (c) shows a significant scatter under synthetic stress amplitude $\mathbf{X}_L^T$ (e.g., points spots in the circle), which is similar to fatigue life scatter under the same stress amplitude mentioned in Figure 4.3.

This work randomly selects 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 4.6 (a) to validate the source task in the MMTL framework for extracting feature embedding from process and defect and classifying defects. This work selects 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 4.6 (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 $\mathbf{X}_L^T \sim \mathrm{U}(0, 0.5)$, as shown in Figure 4.6 (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, this work constructs two graphs in the source task, as shown in Figure 4.7. 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 4.3, potentially improving classification and prediction accuracy.

Table 4.3: 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
$\mathcal{G}_r^S$	250	23421	2	93.68
$\mathcal{G}_2^S$	250	12065	**4**	**48.26**

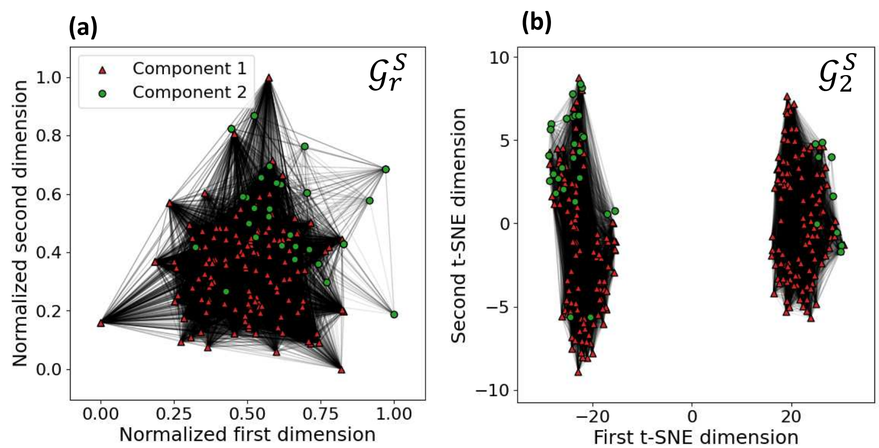

Figure 4.7: The t-distributed Stochastic Neighbor Embedding (t-SNE) visualization of graph representations illustrates graphs (a) $\mathcal{G}_r^S$ and (b) $\mathcal{G}_2^S$ for a synthetic source dataset with 250 data points. The black lines denote the edges between nodes in these graphs. $\mathcal{G}_2^S$ aims to demonstrate that by incorporating feature values and mixture weights, the graph structure can be simplified with fewer edges and a simpler graph structure. Consequently, it becomes easier to differentiate data points compared to $\mathcal{G}_r^S$.

The evaluation of GMM component classification in MMTL source task This work compares HGCN with Hierarchical Neural Network (HNN), Graph Convolutional Network (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 4.4, 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 1.000 on the target dataset, outperforming other benchmark models. This underscores their capability to effectively aggregate useful information from neighboring nodes for accurate classification.

Table 4.4: 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 This work compares 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 the work 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 4.5. 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 4.5: 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	-

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

Defect data description This work assesses the effectiveness of the proposed MMTL in predicting the fatigue life of L-PBF specimens. The details of the experiments are described in Section 4.2.1. This work refers to Set 1 with 1531 volumetric defects as the source data (Figure 4.8 (a)) to train HGCN in MMTL and augmented data with 63 critical defects as the target data (Figure 4.8 (b)) to train NN layers for fatigue life prediction.

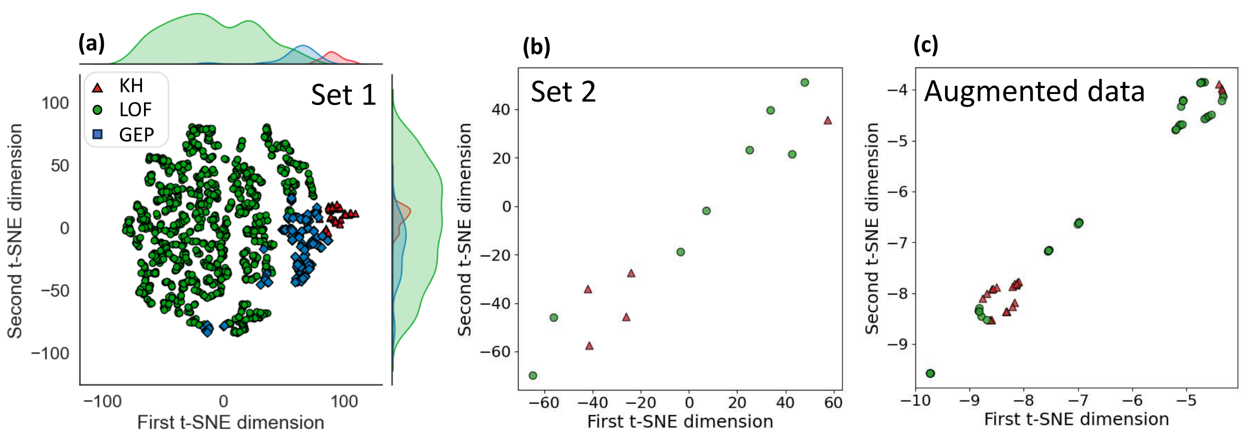

Figure 4.8: The visualization of different types of defects using t-SNE for (a) source data (i.e., Set 1), (b) Set 2, and (c) target data (i.e., augmented data) based on eight defect features mentioned in Table 4.2. The three kernel-estimated distributions on each side in (a) represent the marginal distributions 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).

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

Table 4.6: 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 4.9 (b) that by considering the impact of process parameters on defect features, it 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 4.1 from the graph perspective. Moreover, such integration simplifies graph structure with more concentrated information (i.e., higher average node degree), as shown in Table 4.6, distinguishes defects with unique sizes or morphology, and improves classification and prediction accuracy.

The evaluation of defect classification in MMTL source task This work compares HGCN with HNN, GCN, and NN based on F1-score (full details of model structures in Appendix 8.2.2), as shown in Table 4.7. 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 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

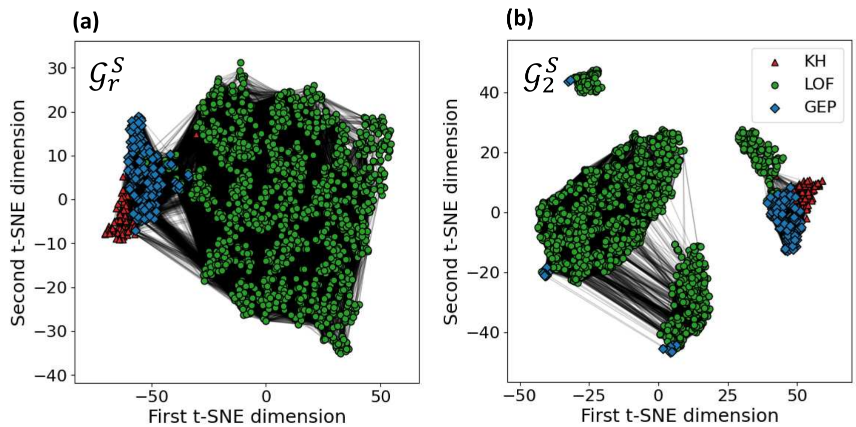

Figure 4.9: The t-SNE graph representation illustrates the graphs (a) $\mathcal{G}_r^S$, (b) $\mathcal{G}_2^S$ for the source dataset. The black lines in these graphs represent the edges between nodes. $\mathcal{G}_2^S$ with both process parameters and defect features has a simpler graph structure and can differentiate defects more easily than $\mathcal{G}_r^S$ with only the defect features.

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 4.7: 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 This work compares the 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 4.4, RMSLE and MAPLE are used in the paper to mitigate the impact of large errors, as the target variable, fatigue life, spans multiple orders of magnitude. In Table 4.8, 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 GCN_2) 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 4.8: 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, this work randomly selects 80% fatigue data to train the NN for fatigue life prediction. Afterward, this work tests 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 [59, 184], as shown in Figure 4.10 (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 4.10 (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 4.9. 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, 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

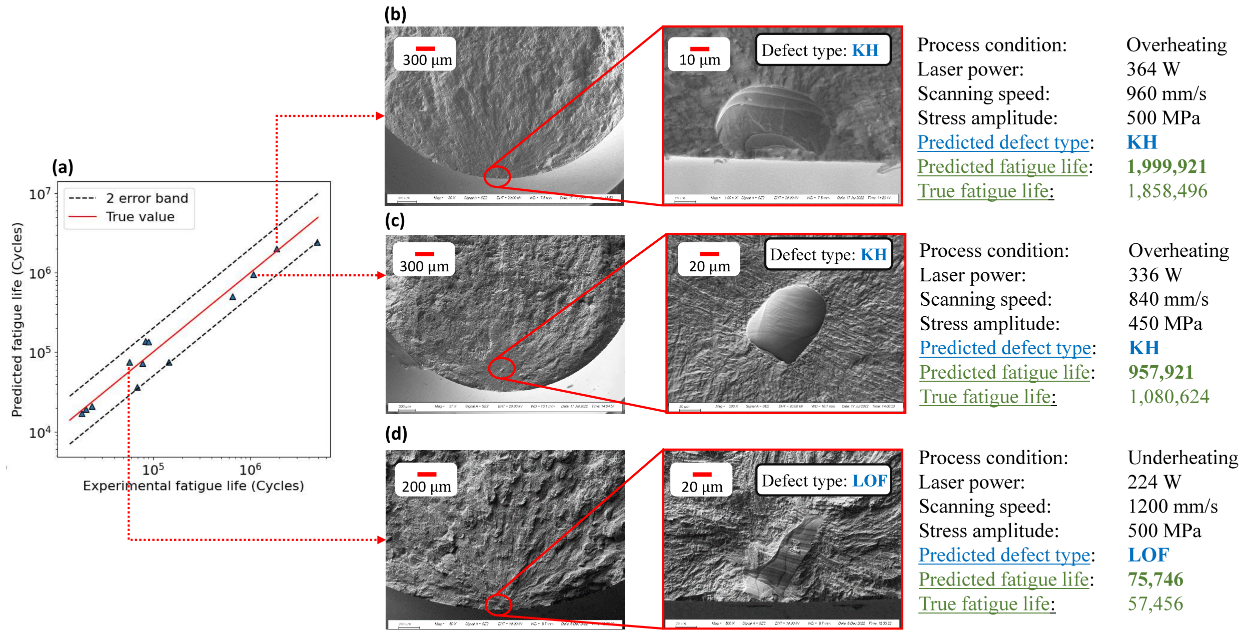

Figure 4.10: (a) The study compares fatigue life predictions between experimentally obtained results and predictions from MMTL. (b-d) represent the critical defect observed in fractography and MMTL prediction results of three distinct L-PBF fatigue testing specimens. MMTL can accurately predict the fatigue life of specimens in (b), (c), and (d).

NN layer for fatigue life prediction effectively balances the bias-variance tradeoff in the 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 4.9: Number of parameters and training time of the compared models in the case study. All models undergo cross-validation to ensure convergence and mitigate overfitting and underfitting.

		Proposed model	Benchmark models			Baseline
Component	**Metric**	**MMTL**	**HNN + NN**	**GCN + NN**	**NN + NN**	**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

4.6 Summary

In this study, an MMTL framework was proposed 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, process and defect data with graphs was represented, leveraging 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 an 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 into 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.

Chapter 5

FedCOT: Personalized Federated Transfer Learning with Conditional Optimal Transport for Manufacturing Predictive Modeling

Effective predictive modeling in large-scale manufacturing is hampered by the isolated and limited data from individual organizations, collected from costly experiments and various inspections. Collaboration across organizations can address these limitations, but it faces two main challenges: privacy concerns over organizations and heterogeneous features from varied sensing and inspection capabilities. Federated learning (FL) offers a solution by allowing organizations to collaboratively train a predictive model without sharing raw data, but standard FL struggles with the problem of feature heterogeneity. To address these challenges, a personalized federated transfer learning framework with conditional optimal transport (FedCOT) is proposed. FedCOT enables ”target” organizations with limited features to benefit from ”source” organizations with sufficient features in prediction performance while keeping data privacy through a central server. Each organization learns a personalized encoder-regressor structure by alternating optimization: the encoder maps heterogeneous inputs into a shared latent space, and the regressor predicts responses from the latent representations. Target organizations align their latent representations' structure and corresponding responses with the source organizations' information via conditional optimal transport. FedCOT is evaluated through simulations and a manufacturing case study on fatigue life prediction of additive-manufactured parts. This case study demonstrates that FedCOT achieves the latent space where target organizations are highly aligned with source organizations, leading to a significant 34.99% improvement in prediction performance over the baseline method. Additionally, this work provides theoretical guarantees on optimal transport gradients, bounded gradients, and predictive consistency.

5.1 Introduction

Effective predictive modeling in manufacturing on a large scale is significantly impeded by the siloed and limited data of individual organizations, obtained from costly experiments and different inspections [1, 185]. Collaboration across organizations is essential to address this limitation, but it faces two key problems: privacy concerns over data sharing [186–188] and heterogeneous features from varied sensing and inspection capabilities. This challenge reflects in manufacturing processes such as laser-powder bed fusion (L-PBF) [11, 189–191], where organizations measure different features due to varied inspection capabilities. For example, understanding fatigue behavior is critical for defect-tolerant design and the structural reliability of manufactured parts [105, 140, 192] by inspecting defects in the fabricated part. While some organizations conduct scanning electron microscope (SEM) for fractography after fatigue testing of fabricated parts for a detailed inspection of defect morphology [193], which is important to identify crack initiation, others only use optical microscopy, which could lose fine defect edges and features due to the lack of a large depth of focus [194]. The different inspection capabilities result in obtaining heterogeneous defect features (e.g., size, shape, and location). Some organizations extract sufficient features for fatigue life prediction, while others are limited to incomplete features. Such a challenging real manufacturing problem, as illustrated in Figure 5.1, motivates the development of a new federated learning (FL) method to improve the prediction performance of organizations with limited inspection capabilities and feature extraction by leveraging knowledge from those with sufficient features without sharing local inspection data.

FL has emerged as a promising solution for collaborative model training to improve prediction performance without sharing raw data [195, 196]. However, standard FL is not sufficient to handle heterogeneous features across organizations, since it prevents FL model aggregation, as shown in Figure 5.1 (a).

In this work, a personalized federated transfer learning framework based on conditional optimal transport (FedCOT) is proposed. The term ”clients” is used to generally refer to ”organizations”, aligning with common FL terminology. The main contributions are summarized as follows:

1. *A personalized federated transfer learning framework for heterogeneous feature spaces.* Under the transfer learning framework, FedCOT enables the ”target” clients with limited features to benefit from the ”source” clients with sufficient features, improving prediction performance while preserving local data privacy.

2. *An encoder-regressor architecture for learning latent representations from heterogeneous features.* Each client adopts an encoder–regressor structure: the encoder maps heterogeneous extracted features into a latent space, and the regressor predicts responses from the latent representations.

3. *Response-aware conditional optimal transport for cross-client latent alignment.* Source clients’ models are pre-trained and aggregated via FedAvg [123] through a central server. Target clients then align their latent features with source representations using response-aware bin-wise conditional optimal transport (COT) and jointly optimize their regressors for personalized prediction via alternating optimization.

4. *Validation through simulations and a real-world additive manufacturing (AM) case study.* This work validates the effectiveness of FedCOT through both simulation studies and a real-world case study on fatigue life prediction in AM.

5.1.1 Research gaps

This work identified three major gaps: (1) existing methods cannot handle clients with heterogeneous features, where some clients lack important predictive features (e.g., defect shape descriptors in AM), (2) current alignment approaches do not properly align the latent space with respect to the response information, which is significant for regression or classification, and (3) most FTL approaches do not focus on multi-client without shared sample settings, whereas real-world manufacturing cases usually require scalable collaboration with strict privacy. The

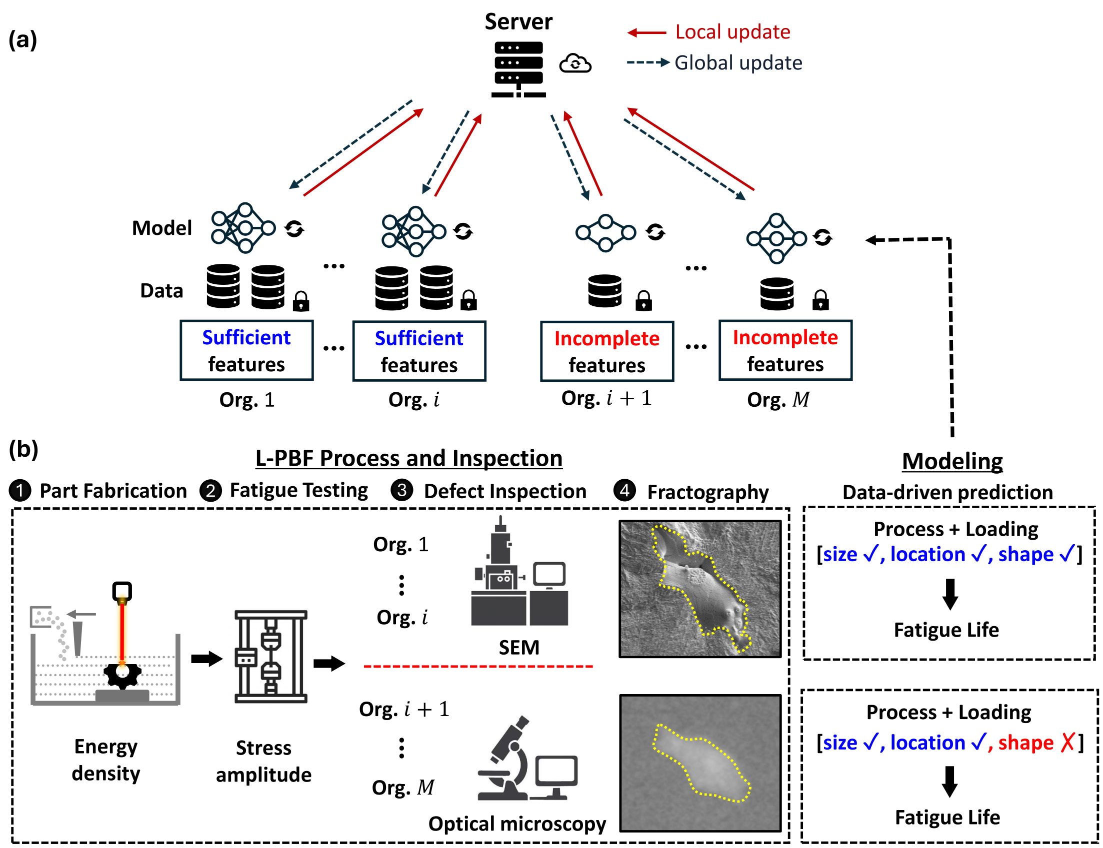

Figure 5.1: The motivation example for the proposed FedCOT framework. (a) Federated learning under heterogeneous feature extraction: Some organizations have sufficient features, and others have incomplete features due to limited inspection capabilities. (b) Illustration of fatigue life prediction in additive manufacturing: (1) laser powder bed fusion process [197], (2) fatigue testing, (3) defect inspection via SEM or optical microscopy, and (4) fractography of defects. High-resolution SEM can capture defect size, location, and shape, while optical microscopy loses fine shape characteristics, leading to incomplete feature sets. "Org." is the abbreviation of "Organization". Note: The image of fractography via optical microscopy shown in this figure is only for demonstration, but does not correspond to actual experimental data.

summary of the literature is shown in Table 2.1. This work develops FedCOT to bridge these gaps in collaborative prediction among clients with limited soloed data and feature heterogeneity while preserving their data privacy.

5.2 Methodology of Personalized Federated Transfer Learning

This work considers a supervised learning setting and assumes there are M clients in total, indexed by $i \in \{1, ..., M\}$. Each client i holds a private dataset $D_i = \{(x_i^{(j)}, y_i^{(j)})\}_{j=1}^{N_i}$, where $x_i^{(j)} \subseteq \mathcal{X}_i \in \mathbb{R}^{d_i}$ is the input vector (e.g., defect features of L-PBF parts for fatigue life prediction), and $y_i^{(j)} \subseteq \mathcal{Y} \in \mathbb{R}$ is a scalar output (e.g., fatigue life of L-PBF parts).

As mentioned in Section 5.1, the key challenge in FL is feature heterogeneity: different clients observe distinct features from the whole feature space $\mathcal{X}$. That is, for each client i, $\mathcal{X}_i \subseteq \mathcal{X}$, and $\mathcal{X}_i \neq \mathcal{X}_h$ for $i \neq h$. As a result, the joint input-output distribution varies across clients $P_i(x, y) \neq P_h(x, y)$ due to:

$$\begin{aligned} &P(x, y) = P(x)\, P(y \mid x), \\ &\text{with } x \sim P(x). \end{aligned} \tag{5.1}$$

where $P_i(x) \neq P_h(x)$ because of different input dimensions across clients and $P_i(y \mid x) \neq P_h(y \mid x)$ due to different parameters of predictive models.

5.2.1 Problem formulation

Based on Eq. (5.1), this work formulates this federated problem as a privacy-preserving transfer learning task. This work assumes there are M_S source clients and M_T target clients with $M_S + M_T = M$. Each source client has sufficient features $x_{S,i} \subseteq \mathcal{X}_S \in \mathbb{R}^{d_o+d_m}$ for accurate prediction, representing real-world organizations with strong inspection capabilities. Here, d_o denotes the number of observed features, and d_m is the number of unobserved (i.e., missing) features. Each target client only observes partial features $x_{T,h} \subseteq \mathcal{X}_T \in \mathbb{R}^{d_o}$, representing real-world small and medium-sized organizations with limited measurement capabilities [198–201].

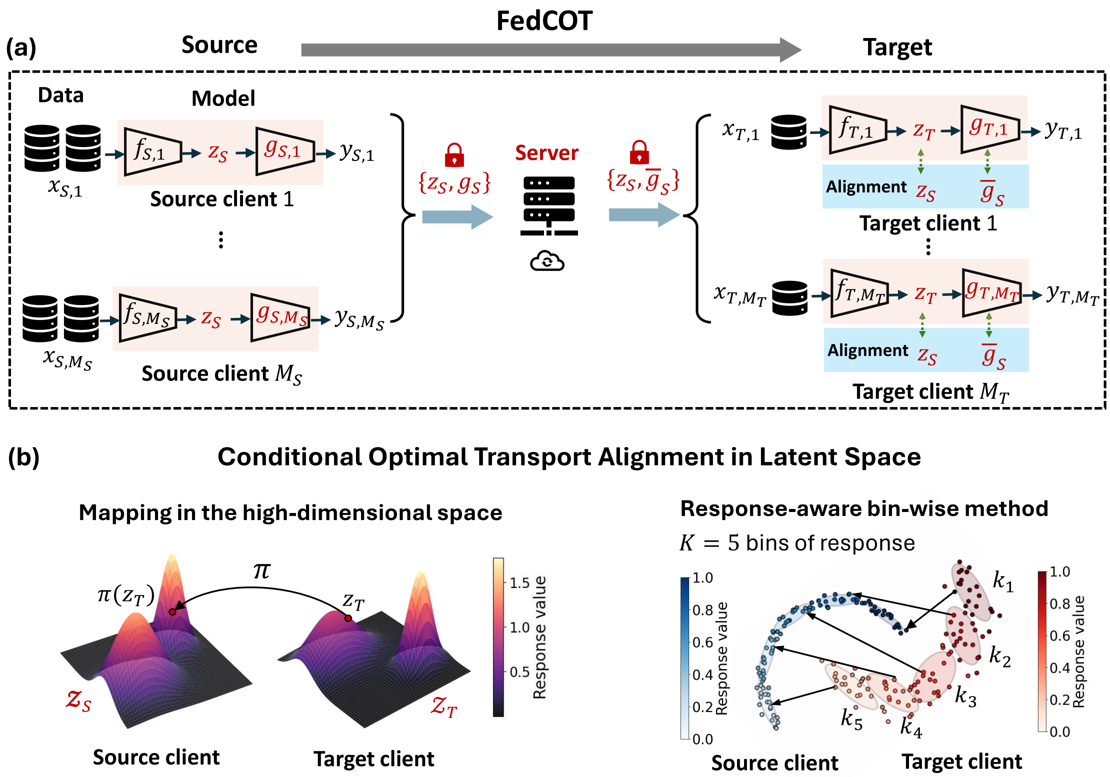

Figure 5.2: Overview of the proposed FedCOT framework. (a) FedCOT structure: the source clients share encoded latent representations and regressors to the server. Then, the target clients align their latent representations with the source clients by conditional optimal transport. (b left) Illustration of distribution alignment in the latent space via optimal transport mapping between source and target manifolds. (b right) An example of quantile-based grouping is in the response space, where ellipses represent quantile bins of data.

The objective is to improve the prediction performance of target clients T by privately transferring knowledge from source clients S.

To address the challenge of feature heterogeneity, this work introduces a latent space $z \in \mathbb{R}^{d_z}$, which maps clients' input data into a compressed latent space and handles different input dimensions across clients. To handle the conditional distribution shift $P(y|x)$ in a private way, for each client, the conditional distribution from Eq. (5.1) can be decomposed as:

$$P(y \mid x) = \int P(z \mid x) P(y \mid z) dz \tag{5.2}$$

Such representation shows the two steps for FL: (1) mapping from input features to latent space, $P(z \mid x)$, and (2) modeling the relationship between latent representations and the response, $P(y \mid x)$. Therefore, in FedCOT, this work adopts an encoder-regressor structure (f_θ, g_β) for all clients. Firstly, the source models $(f_{\theta_S}, g_{\beta_S})$ are pre-trained on the source client dataset. It learns the mapping $f_{\theta_S} : \mathcal{X}_S \to \mathcal{Z}_S$ for $P(z \mid x)$ and the prediction $g_{\beta_S} : \mathcal{Z}_S \to \mathcal{Y}_S$ for $P(y \mid z)$. The learned latent representation z_S and g_S will serve as a reference for target clients to enable knowledge transfer through the central server via FedAvg. Secondly, the target model $(f_{\theta_T}, g_{\beta_T})$ is trained using Alternating Optimization (AO) [202] to optimize the encoder f_{θ_T} and regressor g_{β_T} in an iterative process. At each loop, this work updates the encoder f_{θ_T} by minimizing the Wasserstein distance between z_T and z_S using the proposed bin-wise conditional optimal transport, where latent samples are grouped into different bins based on the similarity of response. This can mitigate the over-matching issue for all data points. Then this work fixes trained f_{θ_T} and updates regressor g_{β_T} with a ridge penalty term to facilitate knowledge sharing from source regressor g_{β_S} (i.e., personalization), which is adopted from [111]. The framework of FedCOT is shown in Figure 5.2 (a).

To guarantee the model's effectiveness, a few key assumptions need to be clearly defined:

Assumption 5.2.1 (Corr-δ assumption). *A federated setting with multiple source and target clients is considered. Each source client S has access to the sufficient input features $x_S \in \mathbb{R}^{d_o + d_m}$, while* *each target client T observes only a partial subset $x_T \in \mathbb{R}^{d_O}$.*

This work assumes that the observed features at the target client are sufficiently correlated with both the missing features (present only at the source) and with the response variable $y \in \mathcal{Y}$. Specifically, there exist constants $\delta_x > 0$ and $\delta_y > 0$ such that

$$|\rho(x_T, x_S)| \geq \delta_x, \qquad |\rho(x_T, y)| \geq \delta_y. \tag{5.3}$$

where $\rho(\cdot, \cdot)$ denotes the Pearson correlation coefficient. In practice, δ_x should be greater than 0.5.

This assumption guarantees two conditions: (i) the target clients' observed features x_T retain predictive relevance for y, and (ii) the target clients' observed features share a non-trivial correlation with the source clients' features x_S, enabling meaningful knowledge transfer. Suppose either correlation falls below the corresponding bound. In that case, the target client features become nearly independent of the response or the missing features, and no knowledge can be shared between the source and target clients.

Importantly, FedCOT is not significantly affected by the number of shared features; instead, its effectiveness depends primarily on whether the observed features retain sufficient conditional information about the missing predictive features. When the target feature set is severely incomplete and the observed features are weakly informative, COT becomes ill-posed. In this case, the target encoder cannot reliably infer missing predictive information, causing latent representations to scatter or collapse toward the mean and to lose predictive structure. As a result, there is no unique transport map from the source latent space to the target latent space, which limits the prediction performance improvement after alignment.

This work also assumes that the response distributions are the same across source and target clients, i.e., $P(y_S) = P(y_T)$, to ensure that the response variable can be an effective and reliable conditioning signal for bin-wise conditional optimal transport (OT).

5.2.2 Conditional optimal transport

This work introduces the core method, bin-wise COT, in FedCOT. This work employs COT to align the latent distributions of target z_T and source z_S in a response-aware bin-wise manner. The goal is to minimize the distributional discrepancy between $P(z_T|y_T)$ and $P(z_S|y_S)$ across different response levels.

To make the conditional alignment tractable, this work discretizes the continuous response y into K equal-mass quantile bins $\{B_k\}_{k=1}^{K}$. Within each bin B_k, this work forms data points $\mathcal{U}_k = \left\{z_T^{(i)} \,:\, y_T^{(i)} \in B_k\right\}, \quad \mathcal{V}_k = \left\{z_S^{(j)} \,:\, y_S^{(j)} \in B_k\right\}$. The cost matrix for bin k is

$$\begin{aligned} C_k^{(ij)} &= \left\| z_T^{(i)} - z_S^{(j)} \right\|_2^2 \\ &= \left\| f_{\theta_T}\left(x_T^{(i)}\right) - f_{\theta_S}\left(x_S^{(j)}\right) \right\|_2^2 \end{aligned} \tag{5.4}$$

This work solves the entropy-regularized optimal transport problem [203]:

$$\pi_k, \theta_T = \arg \min_{\pi \in \Pi(a_k, b_k)} \langle \pi, C_k \rangle + \epsilon \, \mathrm{KL}\left(\pi \| a_k b_k^{\top}\right) \tag{5.5}$$

where a_k and b_k are empirical histograms of $\mathcal{U}_k$ and $\mathcal{V}_k$. The entropy regularization term controls the smoothness of the coupling: small ϵ leads to classical OT, while large ϵ yields softer, more diffuse couplings. And it should always satisfy $\epsilon > 0$ to guarantee a differentiable transport plan.

The overall COT loss is the average over bins:

$$\mathcal{L}_{\mathrm{COT}} = \frac{1}{K} \sum_{k=1}^{K} \mathcal{L}_{\mathrm{OT},k} = \frac{1}{K} \sum_{k=1}^{K} \left(\sum_{i,j} \pi_k^{(ij)} C_k^{(ij)} + \epsilon \sum_{i,j} \mathrm{KL}\left(\pi_k \| a_k^{(i)} b_k^{(j)\top}\right) \right) \tag{5.6}$$

The bin-wise formulation decomposes the fully conditional Wasserstein distance into K tractable sub-problems, enabling efficient optimization and preserving privacy by requiring only the exchange of low-dimensional latent representations rather than raw features. The demonstration is shown in Figure 5.2 (b).

Next, this work justifies the choice of quantile binning rather than other methods like clustering. The quantile binning guarantees balanced sample sizes per bin, which ensures stable empirical distribution matching and tractable optimization. Clustering methods (e.g., k-means in the response space) can introduce instability since they are sensitive to initialization and may cause uneven bin sizes. Also, quantile binning makes it easy to control the quality of the conditional alignment. The number of bins K directly controls its granularity: a small K (e.g., $K = 1$) generates coarse alignment with high bias, whereas a large K leads to sparse bins that make OT unstable and computationally expensive. In practice, it can find the best choice of K via cross-validation.

5.2.3 Training algorithms via alternating optimization with Adam method

So far, this work has formulated the problem and introduced the key alignment method in Eq. (5.5). In this section, this work describes the training and inference procedure of FedCOT via Alternating Optimization with Adaptive Moment Estimation (Adam) under privacy constraints to update both encoder parameters θ and regressor parameters β. The overall FedCOT framework is summarized in Algorithm 1, while the AO is presented in Algorithm 2.

Phase I: Training

Source training Each source client first trains its encoder and regressor by minimizing the objective function:

$$\begin{aligned}\mathcal{L}_S(\theta_S, \beta_S) &= \frac{1}{N_S}\sum_{i=1}^{N_S} \ell\left(g_{\beta_S}(f_{\theta_S}(x_S^{(i)})), y_S^{(i)}\right) \\ &= \frac{1}{N_S}\sum_{i=1}^{N_S} \left(g_{\beta_S}(f_{\theta_S}(x_S^{(i)})) - y_S^{(i)}\right)^2\end{aligned} \tag{5.7}$$

Server (federated averaging) After training, each source client sends pre-trained latent features z_S, regressor g_{β_S}, and the bin indices from B_k to the server. The server will aggregate regressor g_β from all source clients via FedAvg [123]: $\bar{g}_{\beta_S} = \frac{1}{M_S}\sum_{i=1}^{M_S} \beta_{S,i}$. The server then broadcasts global parameters to each target client.

Target training Each target client optimizes their encoder and regressor using the following objective function:

$$\mathcal{L}_T(\theta_T, \beta_T) = \mathcal{L}_{\mathrm{COT}} + \frac{1}{N_T}\sum_{i=1}^{N_T}\left(g_{\beta_T}(f_{\theta_T}(x_T^{(i)})) - y_T^{(i)}\right)^2 + \lambda_\beta \left\|\beta_T - \bar{\beta}_S\right\|^2 \tag{5.8}$$

where $\mathcal{L}_{\mathrm{COT}}$ denotes the bin-wise COT loss, the second term forces prediction accuracy on local data, and the last term is an L_2 ridge penalty to personalize the target client regressor. λ_β controls the level of personalization: a smaller λ_β allows the regressor to keep more individual predictions and more personalization, reducing bias but increasing variance, while a larger one makes the model closer to the source client regressor. In FedCOT, clients are heterogeneous. Some clients benefit from sticking closer to the source (e.g., if their target data is small/noisy). Others benefit from stronger personalization (if their domain is quite different). The hyperparameter explicitly encodes this trade-off.

To solve Eq. (5.8), this work adopts AO shown in Algorithm 2, which contains two steps:

- COT Step: Fix β_T and optimize the encoder θ_T by computing the K bin-wise OT plans $\{\pi_k\}$, aligning target clients' latent features with source clients' latent features. Gradients from $\mathcal{L}_{\mathrm{COT}}$ backpropagate through the cost function to guide the training of the encoder.
- Prediction Step: Fix θ_T and optimize β_T by minimizing $\mathcal{L}_T$ using Adam.

These steps are repeated until convergence.

Phase II: Inference

Once training is finished, each target client can directly obtain predictions for unseen data $\hat{x}_T$ via its personalized encoder-regressor:

$$\hat{y}_T = g_{\hat{\beta}_T}(f_{\hat{\theta}_T}(\hat{x}_T)) \tag{5.9}$$

Algorithm 1: FedCOT Training and Inference

Input: Source datasets $\{D_{S,i}\}_{i=1}^{M_S}$, target datasets $\{D_{T,j}\}_{j=1}^{M_T}$, maximum epoch T_{max}, test inputs $\hat{x}_{T,j}$ for target client j
Output: Predictions $\hat{y}_{T,j}$ for target client j

Phase I: Training

for $t = 1, 2, \ldots, T_{max}$ **do**

```
/* a) Source training */
```

Train each source client i encoder $\theta_{S,i}$ and regressor $\beta_{S,i}$ based on Eq. (5.7);
Each source client i encodes $z_{S,i} \leftarrow f_{\theta_{S,i}}(x_{S,i})$;
Send $z_{S,i}$, local regressor parameters $\beta_{S,i}$, and bin indices from B_k to server;

```
/* b) Server (federated averaging) */
```

Server aggregates via FedAvg:

$$\bar{\beta}_S \leftarrow \text{FedAvg}(\{\beta_{S,i}\}_{i=1}^{M_S})$$

Server distributes $z_{S,i}$, $\bar{\beta}_S$ and bin indices $\{B_k\}$ to each target client j;

```
/* c) Target training */
```

Each target client j performs alternative optimization (Algorithm 2) to update $(\theta_{T,j}, \beta_{T,j})$ locally;

Phase II: Inference

foreach *target client* $j = 1, \ldots, M_T$ **do**

for $v = 1, 2, \ldots, N_v$ **do**

Encode $z_{T,j}^{(v)} \leftarrow f_{\theta_{T,j}}(x_{T,j}^{(v)})$;
Predict $\hat{y}_{T,j}^{(v)} \leftarrow g_{\beta_{T,j}}(z_{T,j}^{(v)})$;

return $\hat{y}_{T,j}^{(v)}$ *for each target client* j

5.2.4 Privacy considerations

In addition to addressing feature heterogeneity, FedCOT also maintains data privacy. First, all training is performed locally, and raw features never leave local clients. Only the source clients' latent representations encoded from input data are shared. Second, this work implements both encoder and regressor as neural networks with nonlinear activations such as ReLU. These mappings are irreversible, and the nonlinear encoder encodes the raw data information into the latent space. Therefore, it is not feasible to recover the raw data from latent features directly. Third, the alignment information is only statistically aggregated to the server and does not

Algorithm 2: Alternating Optimization for Target Client Training

Input: Target dataset $\mathcal{D}_T$, source latent features z_S, global regressor $\bar{\beta}_S$, and bin indices $\{B_S\}$ from server, initialization $\theta_T^{(0)}, \beta_T^{(0)}$

```
/* COT Step: Update encoder θ_T                                    */
```
Fix β_T and update θ_T by minimizing $\mathcal{L}_{\text{COT}}$;

Encode target data: $z_T \leftarrow f_{\theta_T}(x_T)$;

Compute K bin-wise OT plans $\{\pi^{(k)}\}$ between z_T and z_S;

Compute $\mathcal{L}_{\text{COT}}$ and back-propagate to update θ_T;

```
/* Prediction Step: Update regressor β_T                           */
```
Fix θ_T and update β_T by minimizing $\mathcal{L}_T$ in Eq. (5.8) with Adam;

reveal local raw data. Compared to VFL, where shared gradients may still expose partial feature information [204], FedCOT provides additional structural privacy guarantees.

Importantly, FedCOT is easy to equip with existing privacy-preserving methods. Privacy protection methods such as secure aggregation can be integrated on the top layer of the framework to prevent server attacks, and differential privacy (DP) [205] can be incorporated to prevent individual data leakage from model updates. Similarly, homomorphic encryption [206] or trusted execution environments [207] can be used in the framework for high-stakes applications such as healthcare or finance. In summary, FedCOT improves data privacy through its unique encoder-regressor design and provides a flexible framework that can integrate privacy protection methods when required.

5.2.5 Theoretical analysis

This section is about the theoretical insights of the proposed FedCOT framework. This work assumes bounded latent representations ($\|z\|$) and entropic OT regularization with $\epsilon > 0$, which ensures differentiability of the OT loss.

Proposition 5.2.1 (OT Gradient as Barycentric Alignment). *Considering a response bin B_k, target client latent features $U_k = z_T^{(i)}$ and source client latent features $V_k = z_S^{(j)}$, let π_k be the entropic OT plan. Here i and j are sample indices. The gradient of the bin-wise OT loss regarding a target* *latent is*

$$\nabla_{z_T^{(i)}} \mathcal{L}_{\mathrm{OT},k} = 2\left(\sum_j \pi_k^{(ij)} z_T^{(i)} - \sum_j \pi_k^{(ij)} z_S^{(j)}\right) \tag{5.10}$$

Thus, each target latent is pulled toward the barycenter of the source latents under $\pi^{(k)}$. This work focuses on the $\nabla_{z_T^{(i)}} \mathcal{L}_{\mathrm{OT},k}$ since the encoder parameters θ_T are updated via the chain rule for a response bin B_k:

$$\nabla_{\theta_T} \mathcal{L}_{\mathrm{OT}} = \sum_i \frac{\partial \mathcal{L}_{\mathrm{OT}}}{\partial z_T^{(i)}} \cdot \frac{\partial z_T^{(i)}}{\partial \theta_T}. \tag{5.11}$$

The subsequent propagation to θ_T follows standard backpropagation through the encoder architecture.

Moreover, this proposition introduces that the gradient norm is bounded, which introduces a Lemma for the gradient norm:

Lemma 5.2.1 (Bounded OT gradients). *If $\|z_T^{(i)}\|, \|z_S^{(j)}\| \leq G$, then the gradient norm is bounded by*

$$\left\| \nabla_{z_T^{(i)}} \mathcal{L}_{\mathrm{OT},k} \right\| \leq 4G \tag{5.12}$$

Moreover, the Hessian satisfies $\nabla^2_{z_T^{(i)}} \mathcal{L}_{\mathrm{OT},k} = 2w_k^{(i)} I \succeq 0$, where $w_k^{(i)} = \sum_j \pi_k^{(ij)}$, so the OT loss is convex in each latent coordinate.

Proposition 5.2.2 (Predictive Consistency via Conditional OT). *Suppose for each response bin B_k, the conditional distributions of target and source latents satisfy*

$$W_\gamma\left(P\left(z_T \mid y_T \in B_k\right), P\left(z_S \mid y_S \in B_k\right)\right) \leq \gamma \tag{5.13}$$

W_γ is entropic-regularized Wasserstein distance as $W_\gamma(\mu, \nu) = \min_{\pi \in \Pi(\mu,\nu)} \langle \pi, C \rangle + \gamma\, KL(\pi \| \mu \otimes \nu)$. μ and ν are two distributions.

Then the predictive distributions conditioned on response for target and source latents remain close:

$$\sup_{A \subseteq \mathcal{Y}} |P(y_T \in A \mid z_T) - P(y_S \in A \mid z_S)| \leq h(\gamma), \tag{5.14}$$

where $h(\gamma) \to 0$ *as* $\gamma \to 0$*.*

This highlights that the conditional OT ensures the predictive information is preserved when transferring knowledge from the source client to the target client. The full proofs are shown in Appendix A1.

While the FedCOT optimization problem is not convex, the regularizer stabilizes training by biasing the regressor toward the source model and mitigating variance when the target data are limited. By *Lemma 5.2.1 (Bounded OT Gradients)*, the OT term produces stable, Lipschitz-continuous gradients under bounded latent representations and entropic regularization. Consequently, both encoder and regressor optimization satisfy the standard smoothness conditions assumed in nonconvex optimization. These properties ensure that the overall procedure behaves as a block-coordinate descent scheme [208]: the joint loss in Eq. (5.8) decreases monotonically in practice and iterates converge to stationary points. A convergence rate analysis is beyond the scope of this work. Still, FedCOT inherits the typical $\mathcal{O}(1/\sqrt{T})$ sublinear rate to first-order stationary established for nonconvex gradient-based training [111, 124].

5.3 Simulation Study

This section examines the FedCOT performance and robustness by simulated data designed to mimic feature heterogeneity. This work considers two cases: (1) Two-Client Feature-Missing Scenario (Section 5.3.1), where a fully observed source client collaborates with a partially observed target client, representing the case of a resource-rich lab/industry transferring knowledge to a resource-limited lab; and (2) a Multi-Client Feature-Missing Scenario (Section 5.3.2), where multiple target clients observe disjoint subsets of features, mimicking distributed organizations with heterogeneous sensing capabilities and no shared samples. For both studies, FedCOT is assessed against baselines regarding predictive accuracy, error reduction, and alignment performance to show its effectiveness under different degrees of heterogeneity. Since there are no existing methods that can directly tackle the problem, this work designed and implemented several representative baseline models for comparison:

- **Local-only (sufficient features, SF)**: The encoder-regressor model is trained only on the source client data by minimizing the objective function Eq. (5.7). The numerical comparisons for this case were not reported because of the different number of samples compared to FedCOT; instead, this work adopts visualization for qualitative comparison.

- **Local-only (incomplete features, IF)**: The encoder-regressor model is trained only on the target client data. This is also a baseline training without collaboration to compare with FedCOT, and it suffers from missing predictive features and limited data.

- **FedAvg** [123]: A classical federated learning baseline that aggregates local model parameters across clients via weighted averaging. In the setting, FedAvg is implemented by restricting all clients to the overlapping feature subset and training a global encoder-regressor model.

- **FedProx** [95]: An extension of FedAvg that introduces a proximal regularization to constrain local updates toward the global model, aiming to address statistical heterogeneity. Similar to FedAvg, FedProx is implemented on the overlapping features and trains a shared global model.

- **Ditto** [97]: A personalized federated learning method that jointly learns a global model and client-specific local models based on a regularization-based personalization objective. Ditto allows limited personalization and still works on the overlapping features across clients. This work treats Ditto as a representative personalized FL baseline by training both global and local regressors.

- **FedKL**: A federated approach that aligns client features via KL-divergence regularization. This represents a distribution-matching strategy based on probabilistic similarity.

- **FedMMD**: A federated approach that aligns feature distributions using maximum mean discrepancy (MMD). This is a widely adopted kernel-based method for feature alignment.

This work investigates the predictive performance of all benchmark models via root mean square error (RMSE), mean absolute percentage error (MAPE), adjusted R^2, and residual variance. For MAPE, this work adds a small constant (10^{-4}) to the denominator to avoid division by zero. Residual variance can quantify the spread of prediction errors, providing a stability measure of FedCOT under heterogeneous conditions. This work repeats the evaluation ten times to get the mean values and standard deviations to ensure the experiment's reliability. Detailed model settings are in Appendix A2A2.2.1.

5.3.1 Simulation Study 1: Two-Client Feature-Missing Scenario

Setup

In the simulation study, this work simulates one source client and one target client to evaluate the ability of knowledge transfer and prediction improvement: the source client has sufficient input features for prediction, while the target client only has partial observed input features and limited data.

To meet the requirement of the model assumption in Section 5.2.1, this work generates simulated data for the source and target client by sampling the input variables $x = [x_o, x_m]^T \in \mathbb{R}^{d_o+d_m}$ from a multivariate normal distribution, i.e., $x \sim \mathcal{N}(0, \Sigma)$, where x_o are the observed features and x_m are the missing features. The covariance matrix is formulated as

$$\Sigma = \begin{bmatrix} \Sigma_{oo} & \Sigma_{om} \\ \Sigma_{mo} & \Sigma_{mm} \end{bmatrix} \tag{5.15}$$

where $\Sigma_{oo} = (1-\rho)I_{d_o} + \rho\mathbf{1}_{d_o}\mathbf{1}_{d_o}^\top, \quad \Sigma_{mm} = (1-\rho)I_{d_m} + \rho\mathbf{1}_{d_m}\mathbf{1}_{d_m}^\top, \quad \Sigma_{om} =$

$\rho \mathbf{1}_{d_o} \mathbf{1}_{d_m}^\top$. Here, this work assigns $d_o = 3$ and $d_m = 2$. They are the number of observed and missing features, respectively. I is the identity matrix, and $\rho = 0.9$ denotes the pairwise correlation to reflect a strong relationship between observed and missing features.

The regression task is formulated as a nonlinear regression $y = \sigma(w^\top x + b) + \epsilon$, where $\sigma(\cdot)$ denotes a nonlinear activation (e.g., Sigmoid function) and $\epsilon \sim \mathcal{N}(0, \sigma_{\text{noise}}^2)$ models response noise. This mimics the real-world case prediction, where noise arises from stochastic variability and experimental uncertainty. In this simulation, this work assigns the same weights for each input feature as $w = [1, 1, ..., 1]^T \in \mathbb{R}^5$, $b = 0.5$, and $\sigma_{\text{noise}} = 0.01$. This work generates $N_{\text{source}} = 200$ samples for the source client and $N_{\text{target}} = 100$ samples for the target client, reflecting the practical scenario where the target client has fewer data points than the source. Also, this work defines the encoder and regressor in FedCOT as neural networks. Ideally, the encoder can learn to project heterogeneous input features into the shared latent space aligned with the source client, while the regressor can learn the predictive mapping from latent representations to the response. The proportion of the test set is 20%.

Simulation results

Table 5.2 reports prediction performance when one source client collaborates with three target clients with heterogeneous features. FedCOT achieves the best performance across all metrics compared to Local-only (incomplete features), FedAvg, FedProx, Ditto, FedKL, and Fed-MMD, with RMSE of 0.0728 and MAPE of 0.0146. Compared to the Local-only (incomplete features), FedCOT improves RMSE by 37.83% and MAPE by 54.68%, while also achieving a higher adjusted R^2 (0.9628) and lower residual variance (0.0050). This indicates that FedCOT preserves the predictive structure by regressor across clients with heterogeneous inputs and shows more stable predictions with low residual variance. These results demonstrate that FedCOT scales effectively to multi-client settings and can leverage source knowledge to improve target clients' prediction performance, even when some target clients lack critical predictive features.

Figure 5.3 and Figure 5.4 further support the findings. In the latent space, it is noted that

Table 5.1: Simulation performance for the prediction of the Multi-Client scenario. Note: The results are averaged across Clients A, B, and C. Reported values are mean ± std over 10 runs. "IF" indicates "Incomplete Features".

Method	RMSE	MAPE	Adj. R^2	Residual Var.
Local-only (IF)	0.1234 (0.0152)	0.0681 (0.1566)	0.8995 (0.0277)	0.0145 (0.0038)
FedAvg	0.1161 (0.0105)	0.0545 (0.0085)	0.9133 (0.0166)	0.0131 (0.0025)
FedProx	0.1102 (0.0058)	0.0942 (0.0331)	0.9213 (0.0066)	0.0116 (0.0016)
Ditto	0.0957 (0.0091)	0.0573 (0.0095)	0.9384 (0.0155)	0.0086 (0.0016)
FedKL	0.0952 (0.0115)	0.0805 (0.1752)	0.9413 (0.0160)	0.0084 (0.0021)
FedMMD	0.1179 (0.0205)	0.0827 (0.2215)	0.9043 (0.0332)	0.0130 (0.0039)
FedCOT (Ours)	**0.0728** **(0.0158)**	**0.0174** **(0.0465)**	**0.9628** **(0.0167)**	**0.0050** **(0.0024)**
Gain (%)	↑ 37.83	↑ 54.68	↑ 6.36	↑ 60.31

FedCOT produces a more compact and well-aligned embedding, with response distributions (color gradient in the figure) that closely match the Local-only (sufficient features) model, unlike Local-only (incomplete features), FedAvg, FedProx, Ditto, FedKL, or FedMMD, as shown in Figure 5.3. In the prediction scatter plots, FedCOT demonstrates tighter prediction around the ideal line (Figure 5.4 (c)), indicating not only lower error but also more stable prediction. These results illustrate that FedCOT can transfer knowledge from the source client to the target client while preserving the client's private predictive structure, enabling accurate alignment and reliable regression performance under feature heterogeneity.

5.3.2 Simulation Study 2: Multi-Client Feature-Missing Scenario

Setup

This work uses the same data generation procedure as in Simulation Study 1, but extends it to one source client and three target clients to evaluate the scalability under heterogeneous feature spaces. In this study, the source client retains all five input features, while the target clients observe only partial features: Client A with $x_{\mathrm{A}} = [x_1, x_2, x_3]^T$, Client B with $x_{\mathrm{B}} = [x_2, x_4, x_5]^T$,

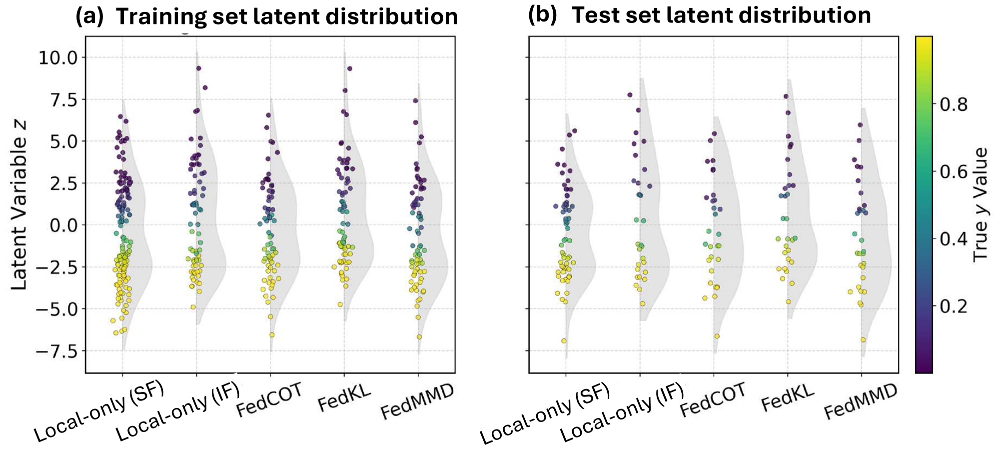

Figure 5.3: Comparison of latent space distributions for (1) training and (b) test sets under Local-only (sufficient features), Local-only (incomplete features), FedCOT, FedKL, and FedMMD. FedCOT produces more consistent latent representations with the Local-only (sufficient features) model regarding structure as well as y distribution, indicating effective knowledge transfer.

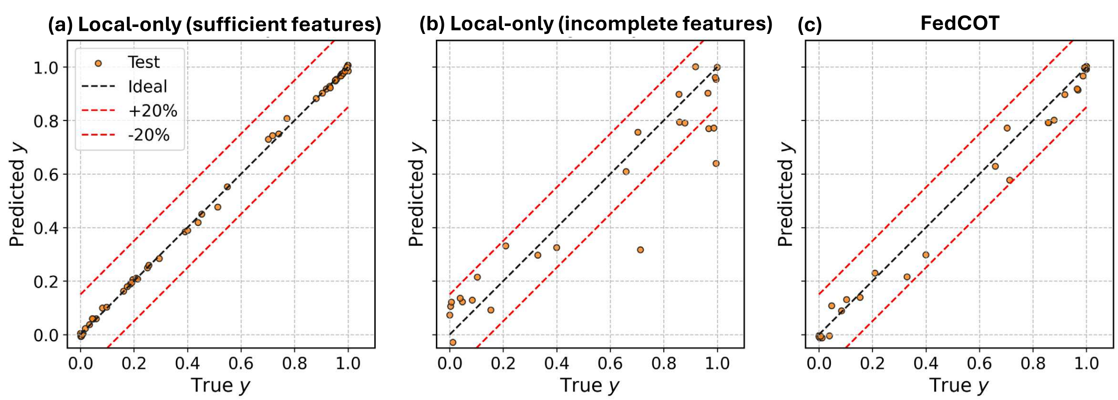

Figure 5.4: The prediction performance of predicted vs. true responses for (a) Local-only (sufficient features), (b) Local-only (incomplete features), and (c) FedCOT on the test set. FedCOT demonstrates more accurate predictions than the Local-only (incomplete features) model, indicating its ability to leverage source knowledge to improve target client prediction performance.

and Client C with $x_C = [x_4, x_5]^T$. This simulation reflects realistic federated scenarios where different clients collect distinct features depending on their sensing or testing capabilities, and in some cases (e.g., Client C) have severely limited information.

Simulation results

Table 5.2 reports prediction performance when one source client collaborates with three target clients with heterogeneous features. FedCOT achieves the best performance across all metrics compared to Local-only (incomplete features), FedAvg, FedProx, Ditto, FedKL, and FedMMD, with RMSE of 0.0728 and MAPE of 0.0146. Compared to the Local-only (incomplete features), FedCOT improves RMSE by 37.83% and MAPE by 54.68%, while also achieving a higher adjusted R^2 (0.9628) and lower residual variance (0.0050). This indicates that FedCOT preserves the predictive structure by regressor across clients with heterogeneous inputs and shows more stable predictions with low residual variance. These results demonstrate that FedCOT scales effectively to multi-client settings and can leverage source knowledge to improve target clients' prediction performance, even when some target clients lack critical predictive features.

Table 5.2: Simulation performance for the prediction of the Multi-Client scenario. Note: The results are averaged across Clients A, B, and C. Reported values are mean ± std over 10 runs. "IF" indicates "Incomplete Features".

Method	RMSE	MAPE	Adj. R^2	Residual Var.
Local-only (IF)	0.1234 (0.0152)	0.0681 (0.1566)	0.8995 (0.0277)	0.0145 (0.0038)
FedAvg	0.1161 (0.0105)	0.0545 (0.0085)	0.9133 (0.0166)	0.0131 (0.0025)
FedProx	0.1102 (0.0058)	0.0942 (0.0331)	0.9213 (0.0066)	0.0116 (0.0016)
Ditto	0.0957 (0.0091)	0.0573 (0.0095)	0.9384 (0.0155)	0.0086 (0.0016)
FedKL	0.0952 (0.0115)	0.0805 (0.1752)	0.9413 (0.0160)	0.0084 (0.0021)
FedMMD	0.1179 (0.0205)	0.0827 (0.2215)	0.9043 (0.0332)	0.0130 (0.0039)
FedCOT (Ours)	**0.0728 (0.0158)**	**0.0174 (0.0465)**	**0.9628 (0.0167)**	**0.0050 (0.0024)**
Gain (%)	↑ 37.83	↑ 54.68	↑ 6.36	↑ 60.31

Figure 5.5 shows the prediction scatter plot for all target clients regarding Local-only (incomplete features) and FedCOT. Across all three target clients, FedCOT yields predictions that more points around the ideal line within the ±20% error bounds, confirming the error reductions in RMSE and MAPE reported in Table 5.2. Notably, even Client C - observing only two features - still achieved reliable predictions through conditional alignment with the source client. These results support that FedCOT improves overall prediction and preserves predictive consistency across clients, validating its scalability under heterogeneous features.

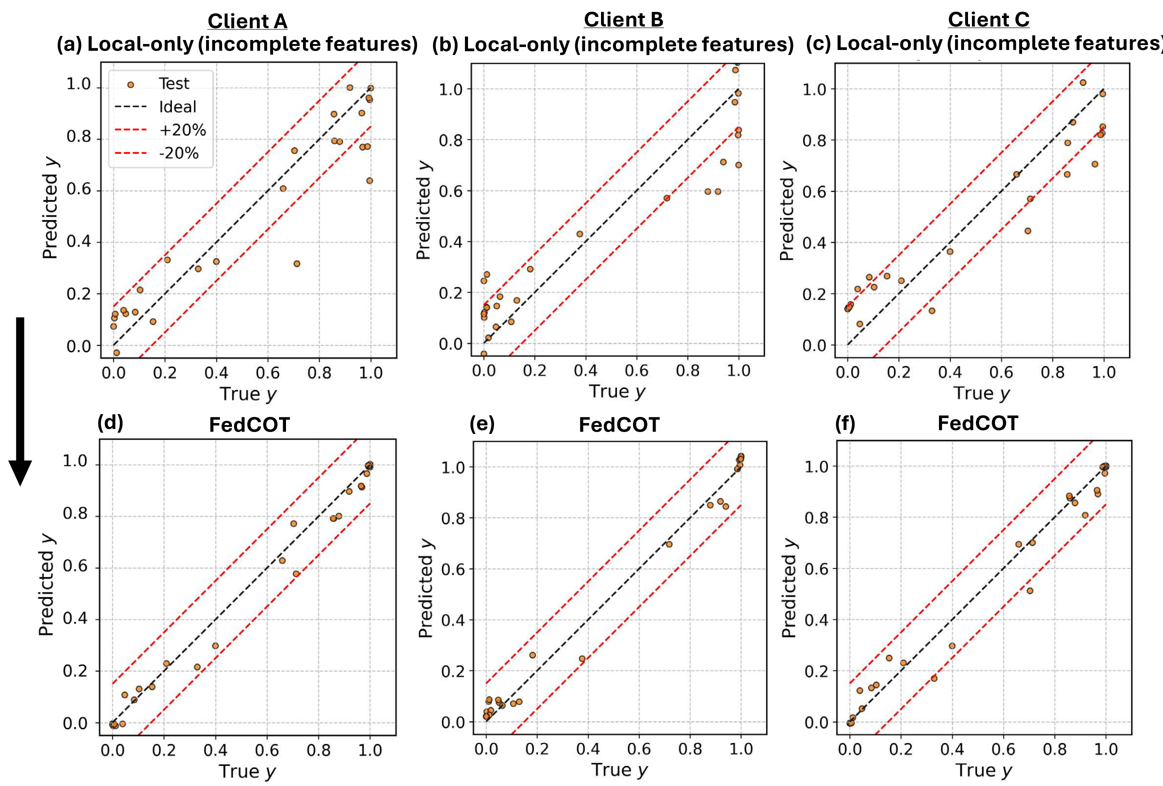

Figure 5.5: The prediction performance of predicted vs. true responses for (a)-(c) Local-only (incomplete features), and (d)-(f) FedCOT on test sets across different clients. FedCOT demonstrates more accurate predictions.

5.3.3 Ablation and Sensitivity Analysis

To evaluate the design of the proposed FedCOT framework and investigate its robustness, this work conducts an ablation study and sensitivity analysis. Specifically, this work examines (1) the contribution of each key FedCOT component, and (2) the impact of data availability and

Table 5.3: Ablation study of key components in FedCOT. This table evaluates the contribution of each key component of the proposed FedCOT framework: OT, bin-wise conditional alignment, ridge regularization, and alternating optimization. Reported values are mean ± std over 10 runs. ”Ridge Reg.” indicates ”ridge regularization”, and ”Alternating Opt.” indicates ”alternating optimization”.

Ablation Settings				**Model Performance**			
OT	Bin-wise (K=5)	Ridge Reg.	Alternating Opt.	RMSE ↓	MAPE ↓	Adj. R^2 ↑	Residual Var. ↓
×	×	×	×	0.1029 (0.0093)	0.1245 (0.2733)	0.9299 (0.0175)	0.0099 (0.0022)
✓	×	×	×	0.0897 (0.0061)	0.0434 (0.2733)	0.9476 (0.0081)	0.0078 (0.0011)
✓	✓	×	×	0.0793 (0.0082)	0.0427 (0.0495)	0.9591 (0.0073)	0.0059 (0.0016)
✓	✓	✓	×	0.0769 (0.0120)	0.0206 (0.0372)	0.9602 (0.0084)	0.0056 (0.0013)
✓	✓	✓	✓	**0.0728** (0.0158)	**0.0174** (0.0465)	**0.9628** (0.0167)	**0.0050** (0.0024)

critical hyperparameters.

Ablation study of FedCOT components

This work performs an ablation study to justify the significance of each key component in FedCOT, including OT, bin-wise conditional alignment, ridge regularization for the target regressor, and alternating optimization. The results are shown in Table 5.3.

Without the proposed components, the model exhibits limited prediction performance, as the source client cannot effectively transfer knowledge to the target clients across heterogeneous feature spaces. Adding OT alignment enhances prediction accuracy and reduces variance, demonstrating the benefits of distribution-level alignment in the latent space. Further introducing bin-wise manner (with $K = 5$) consistently improves performance, indicating that response-aware bin-wise conditioning alignment mitigates regression heterogeneity between source and target clients.

Incorporating ridge regularization (i.e., personalization) to the target regressor stabilizes the regression model and prevents overfitting to noisy or limited target data. Finally, adding alternating optimization (i.e., FedCOT) achieves the best performance. This result emphasizes the importance of each component to the FedCOT in terms of balancing bias and variance during personalization.

Sensitivity to feature correlation

To evaluate the robustness of FedCOT, this work conducts a sensitivity analysis on the correlation structure among features. This experiment demonstrates the FedCOT performance with varying feature correlation, which can provide analysis for the Assumption 3.1 (*Corr-δ assumption*) in Section 5.2.

To assess the efficiency of knowledge transfer based on different feature correlations, this work generates input variables $x \sim \mathcal{N}(0, \Sigma)$ under different correlation levels $\rho \in \{0.3, 0.5, 0.7, 0.9\}$, where Σ has ones on the diagonal and ρ on the off-diagonal. This tests whether FedCOT can better leverage correlated, but inaccessible source client features to transfer to the target client.

Table 5.4 examines FedCOT under different feature correlation levels ρ. Its performance improves as correlation increases, with adjusted R^2 increasing from 0.6352 ($\rho = 0.3$) to 0.9676 ($\rho = 0.9$). The gains over the Local-only (incomplete features) model also grow with correlation, reaching a 37.43% reduction in RMSE when $\rho = 0.9$. This highlights the key property of FedCOT: by conditioning alignment on the response, the model can effectively leverage correlated feature information from the source to the target client. Knowledge transfer is naturally more limited in low-correlation settings, yet FedCOT still provides benefits over baselines. It confirms that FedCOT is most effective when predictive features show strong correlations.

Sensitivity to source sample size

This work evaluates the impact of data availability at the source client by analyzing target prediction performance as the number of source samples changes. Figure 5.6 shows RMSE under different source samples while keeping the target data fixed. The local-only (IF) baseline does not leverage any source information. The performance of all FL methods improves with an increased

Table 5.4: The results of FedCOT sensitivity to correlation strength ρ in the covariance matrix. Reported values are mean $\pm$ std over 10 runs.

ρ	RMSE	MAPE	Adj. R^2	Residual Var.	Gain (%)
0.3	0.2226 (0.0301)	0.0324 (0.0583)	0.6352 (0.1122)	0.0469 (0.0137)	↑ **9.38**
0.5	0.1680 (0.0398)	0.0246 (0.0325)	0.7946 (0.0842)	0.0270 (0.0128)	↑ **12.72**
0.7	0.1246 (0.0360)	0.0198 (0.0201)	0.8874 (0.0584)	0.0147 (0.0088)	↑ **26.95**
0.9	0.0678 (0.0148)	0.0172 (0.0398)	0.9676 (0.0179)	0.0042 (0.0022)	↑ **37.43**

number of source samples. Notably, FedCOT consistently achieves the lowest RMSE across all source sample sizes and shows significant gains when the source client has fewer data points (e.g., 50 samples). This result suggests that FedCOT is better suited to situations where source and target data are limited, which is common in real-world manufacturing applications.

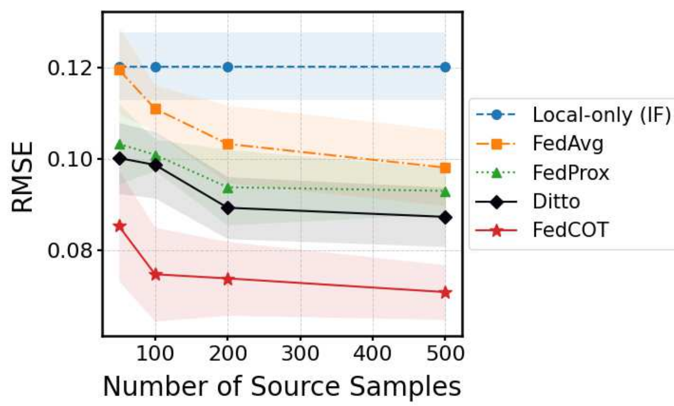

Figure 5.6: The impact of the source sample size on target prediction performance. Local-only (IF) is the baseline model; FedAvg, FedProx, and Ditto show error decreases with increasing source samples. Notably, FedCOT consistently achieves the lowest RMSE across all source sample sizes. The error band denotes one standard deviation over 10 runs.

Sensitivity to conditional binning parameter K

This work next investigates the sensitivity of FedCOT to the number of bins K used in COT alignment. Table 5.5 exhibits the RMSE under different values of K.

It is noted that the performance is poor when $K = 1$ since the method becomes unconditional optimal transport for distribution-level alignment. Increasing K improves performance up to $K = 5$, which achieves the lowest RMSE. However, further increasing K leads to poor performance due to insufficient samples per bin and increased alignment variance. These results suggest that an intermediate binning level provides a favorable trade-off between alignment performance and statistical stability.

Sensitivity to ridge regularization parameter λ_β

Finally, this work examines the sensitivity of FedCOT to the ridge regularization parameter λ_β, as shown in Table 5.5. The regressor tends to overfit to the limited target data when $\lambda_\beta = 0$, resulting in a poor performance. Selecting moderate regularization significantly reduces RMSE, with the best performance achieved at $\lambda_\beta = 0.5$. In contrast, large regularization values degrade performance by over-constraining personalization.

Table 5.5: Sensitivity analysis of FedCOT considering the number of bins K and ridge penalty λ_β. Reported values are mean $\pm$ std over 10 runs.

Hyperparameter	Value	RMSE
K (Number of bins)	1	0.0959 (0.0088)
	3	0.0903 (0.0059)
	5	**0.0738** (0.0060)
	7	0.0880 (0.0054)
	10	0.0902 (0.0059)
λ_β (Ridge penalty)	0	0.0943 (0.0097)
	0.01	0.0908 (0.0065)
	0.5	**0.0738** (0.0080)
	1	0.1030 (0.0083)
	10	0.0914 (0.0058)

5.4 Case Study of Distributed Fatigue Life Prediction of L-PBF Parts under Feature Heterogeneity

In this section, this work focuses on the evaluation of FedCOT on a real-world case study - fatigue life prediction of L-PBF parts in the Federated setting. This work will discuss the problem description in Section 5.4.1, the experimental setup in Section 5.4.2, and the results in Section 5.4.3.

5.4.1 Problem description

As mentioned in Section 5.1, fatigue life prediction is crucial for ensuring structural reliability and qualification of AM parts. In L-PBF, defects such as lack-of-fusion pores or keyholes serve as a crack initiation site and affect fatigue performance [105, 209]. In the realistic industrial scenario, however, not all manufacturers have access to expensive equipment such as SEM for detailed defect inspection. For example, SEM provides finer and detailed morphology information for defects (e.g., circularity, aspect ratio, or solidity) at micron-scale resolution (0.5-2 μm per pixel), but it requires a costly setup [210]. In contrast, optical microscopy is faster and more accessible, but only captures coarse features such as general defect size and location. This reflects the resource gap between large manufacturers equipped with expensive inspection capabilities and small-and medium-sized manufacturers (SMMs) that rely on affordable inspection methods. This motivates the case study design: source clients as well-equipped manufacturers with sufficient defect features for fatigue life prediction, while target clients represent SMMs limited to optical microscopy. FedCOT is designed to improve target clients' fatigue life prediction performance by leveraging the defect shape knowledge from source clients without sharing the inspection data.

5.4.2 Experiment setup and federated data setup

L-PBF Metal Parts Fabrication and Fatigue Testing

The dataset used in this study consists of 156 Ti-6Al-4V cylindrical rod specimens fabricated by laser powder bed fusion (L-PBF) on an EOS M290 system using plasma-atomized Grade 5 powder. The dataset was provided by the National Center for Additive Manufacturing Excellence (NCAME). Eleven process conditions of the energy density of the laser from 42.07 - 95.24 J/mm^3 were investigated, spanning lack-of-fusion to keyhole defect regimes. In detail, all rods were post-processed by stress-relieving, machining, and polishing into tensile and fatigue specimens

(ASTM E08 [211] and ASTM E466 [212]). Fatigue tests were conducted at room temperature under fully reversed loading ($R_\sigma = -1$) at 450, 500, 600, and 700 MPa controlled stress amplitudes. At least seven specimens per stress amplitude were tested (excluding run-outs at 10^7 cycles). After that, fractography was carried out using SEM to identify the critical defect (i.e., crack-initiating defect) that causes the failure.

The final dataset consists of eight predictive features and one response variable: two process/loading features (energy density, stress amplitude), five critical defect features (the distance of the defect to the surface, Murakami size [213], circularity, aspect ratio, and solidity), and the response variable, fatigue life (cycles to failure). The shape-related features were extracted by a computer vision content-based image retrieval (CBIR) method based on the binarized defect contour [11].

Federated data setup

- **Two-client federated setting**: This work first considers one source client and one target client to evaluate FedCOT performance. The source client has the complete dataset, containing all 156 samples with eight input features and the response of fatigue life. The target client was constructed by randomly sampling 100 samples (features and response pairs) from the complete dataset with replacement, then adding Gaussian noise to simulate measurement uncertainty. Specifically, for each input feature and the response, this work added noise drawn from $\mathcal{N}(0, 0.05\sigma_j^2)$, where σ_j is the empirical standard deviation of the j-th input feature. Then, the target client only has energy density, stress amplitude, the distance of the defect to the surface, defect size, and fatigue life. This boostrapping with noise procedure mimics realistic industrial conditions where the target client has limited and different samples due to time-consuming fatigue testing, compared to the source client.

- **Multi-client federated setting**: Building on the two-client case study, this work further extends the study to a multi-client FL setting involving one source client and three heterogeneous target clients. All target clients' data are constructed in the same bootstrapping and noise-injection way as in the two-client setting. Clients with different levels of heterogeneity exhibit varying inspection and characterization capabilities. Client A has no defect morphology features, Client B lacks defect size and morphology, and Client C is the most constrained setting and has no access to defect-related features. The summary of clients' information is shown in Table 5.6. This multi-client setting evaluates fatigue life prediction under increasing levels of feature heterogeneity and assesses the robustness and scalability of FedCOT when multiple data-limited, less-equipped manufacturers participate in FL. The full details of the model architecture are shown in Appendix A2A2.2.2.

Table 5.6: Client feature availability and severity in the multi-client federated setting.

Client	**Severity (Missing Information)**
Source client	None
Client A	Mild (partial defect features missing)
Client B	Moderate (limited defect information)
Client C	Severe (no defect information)

5.4.3 Results

Two-client case study

Table 5.7 shows the prediction performance of FedCOT compared with baseline models on the real fatigue life dataset. FedCOT outperforms other models, achieving RMSE of 0.3104 and MAPE of 0.0490, representing 34.9% and 32.2% improvements over the Local-only (incomplete features) model, respectively, while also significantly increasing adjusted R^2 and reducing residual variance. It indicates that COT alignment enables effective knowledge transfer in terms of structural and response information from the source to the target client.

Figure 5.8 provides further evidence: FedCOT produces a compact and well-aligned embedding similar to the Local-only (sufficient features) latent space in both structure and response distribution (color gradient). The alignment success comes from two factors: (1) the

Table 5.7: Prediction performance comparison on the two-client case study dataset. FedCOT consistently achieves the lowest error compared to Local-only (incomplete features), FedAvg, FedProx, Ditto, FedKL, and FedMMD. Reported values are mean ± std over 10 runs. "IF" indicates "Incomplete Features". The value of "-" means an invalid negative R^2.

Method	**RMSE**	**MAPE**	**Adj. R^2**	**Residual Var.**
Local-only (IF)	0.4821 (0.0540)	0.0713 (0.0106)	0.2209 (0.1826)	0.2642 (0.1797)
FedAvg	0.4047 (0.0158)	0.0618 (0.0038)	0.4813 (0.1148)	0.1458 (0.0245)
FedProx	0.3896 (0.0227)	0.0601 (0.0064)	0.5245 (0.0996)	0.1414 (0.0186)
Ditto	0.3564 (0.0127)	0.0539 (0.0028)	0.6047 (0.0828)	0.0986 (0.0265)
FedKL	0.6622 (0.1029)	0.1054 (0.0151)	-	0.4135 (0.1383)
FedMMD	0.6564 (0.0582)	0.1020 (0.0132)	-	0.4107 (0.0690)
FedCOT (Ours)	**0.3104 (0.0205)**	**0.0470 (0.0053)**	**0.6942 (0.0849)**	**0.0927 (0.0159)**
Gain (%)	↑ 34.99	↑ 32.18	↑ 182.54	↑ 58.19

response-aware bin-wise design of COT, and (2) the statistical correlations between missing defect morphology features (circularity, aspect ratio, and solidity) and observed features (e.g., energy density and defect size), as shown in Figure 5.7. These correlations support the assumption in Section 5.2.1 that missing features are not independent of observed features, enabling feasible and effective knowledge transfer.

Finally, Figure 5.9 demonstrates that FedCOT predictions follow the ideal line more closely than the Local-only (incomplete features) model, with fewer large deviations and improved generalization on both the training and test sets. It confirmed that COT improves mapping in latent space and translates into a more accurate and stable fatigue life prediction with the following personalized regressor. These results showed that FedCOT effectively integrates heterogeneous features, preserves predictive power, and is applicable to real-world fatigue life prediction with limited and noisy data.

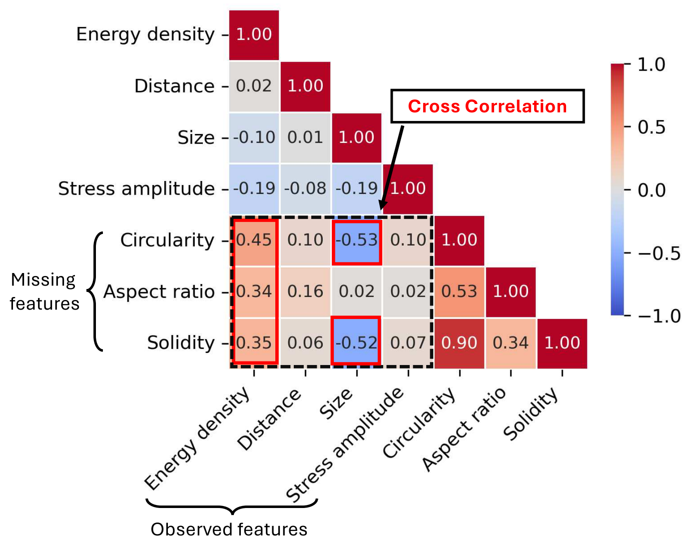

Figure 5.7: Correlation heatmap of observed and missing features. Observed features (*Energy density, Distance, Size, Stress amplitude*) are grouped at the bottom, and missing defect shape-related features (*Circularity, Aspect ratio, Solidity*) are grouped on the left. The dashed black box denotes the missing feature block, and highlighted cells illustrate their correlations with observed features.

Multi-client case study

Table 5.8 shows the fatigue life prediction performance of different methods in the multi-client case study, where one source client collaborates with three target clients with increasing levels of feature missingness. FedCOT consistently achieves the best performance compared to Local-only (IF), FedAvg, FedProx, Ditto, FedKL, and FedMMD. In particular, FedCOT obtains the lowest RMSE (0.3971) and MAPE (0.0601). Compared to the Local-only baseline, FedCOT enhances RMSE by 34.57% and MAPE by 21.7%, demonstrating a significant reduction in prediction error.

These results indicate that FedCOT effectively preserves the predictive structure of the regressor for fatigue life prediction across target clients with heterogeneous, incomplete input features. In contrast, baseline methods such as FedAvg, FedProx, and Ditto show limited improvement over Local-only (IF), while distribution alignment-based methods (FedKL and FedMMD) perform poorly due to a lack of conditional alignment between latent representations and fatigue responses.

Figure 5.10 further demonstrates the prediction accuracy across individual target clients

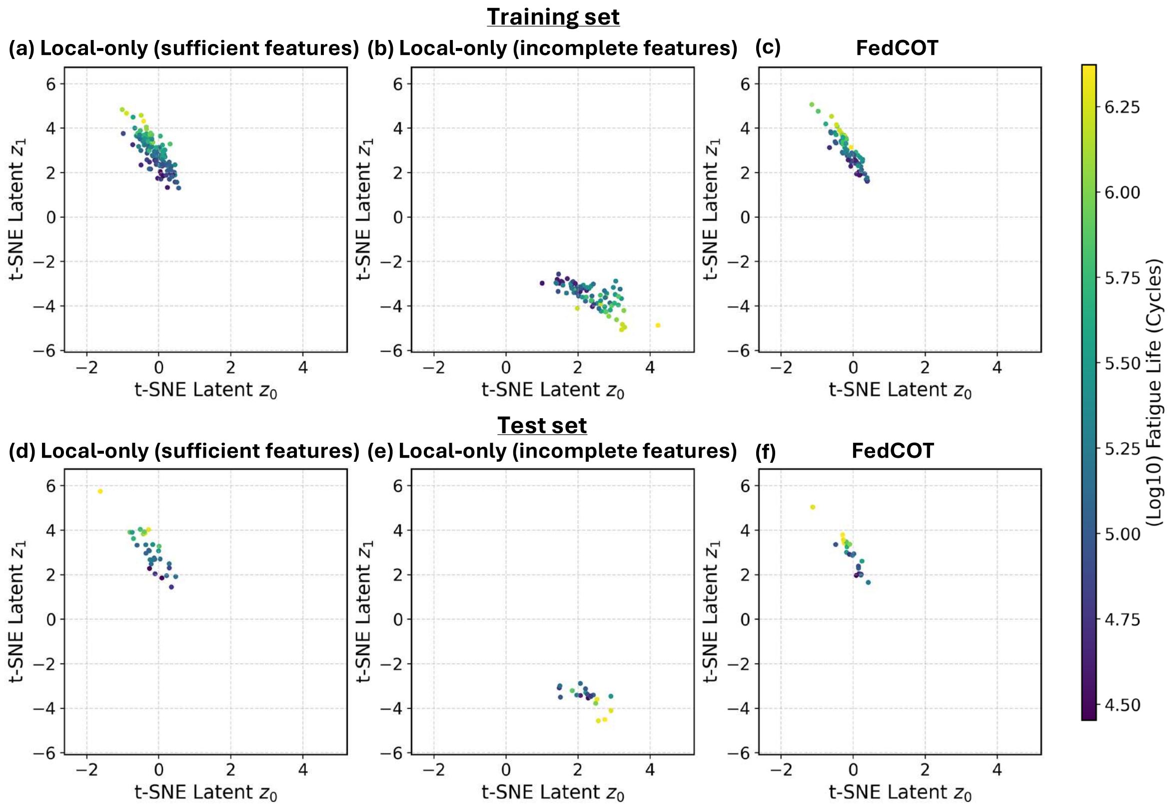

Figure 5.8: The t-distributed stochastic neighbor embedding (t-SNE) visualization of latent representations for (a) and (d) Local-only (sufficient features), (b) and (e) Local-only (incomplete features), and (c) and (f) FedCOT on training and test sets. FedCOT produces more structured and aligned latent spaces with Local-only (sufficient features), indicating improved conditional alignment.

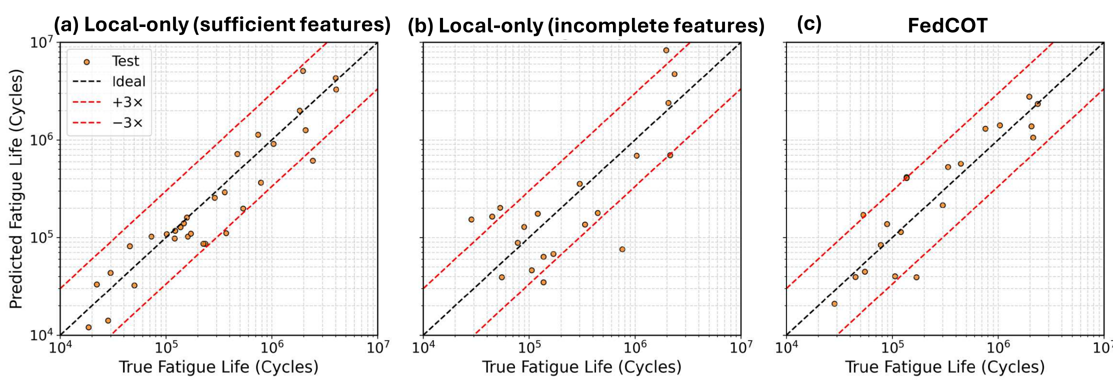

Figure 5.9: The prediction performance of predicted vs. true fatigue life for (a) Local-only (sufficient features), (b) Local-only (incomplete features), and (c) FedCOT on test sets. The diagonal line indicates ideal prediction. FedCOT yields more accurate and consistent predictions than Local-only (incomplete features), and is close to Local-only (sufficient features).

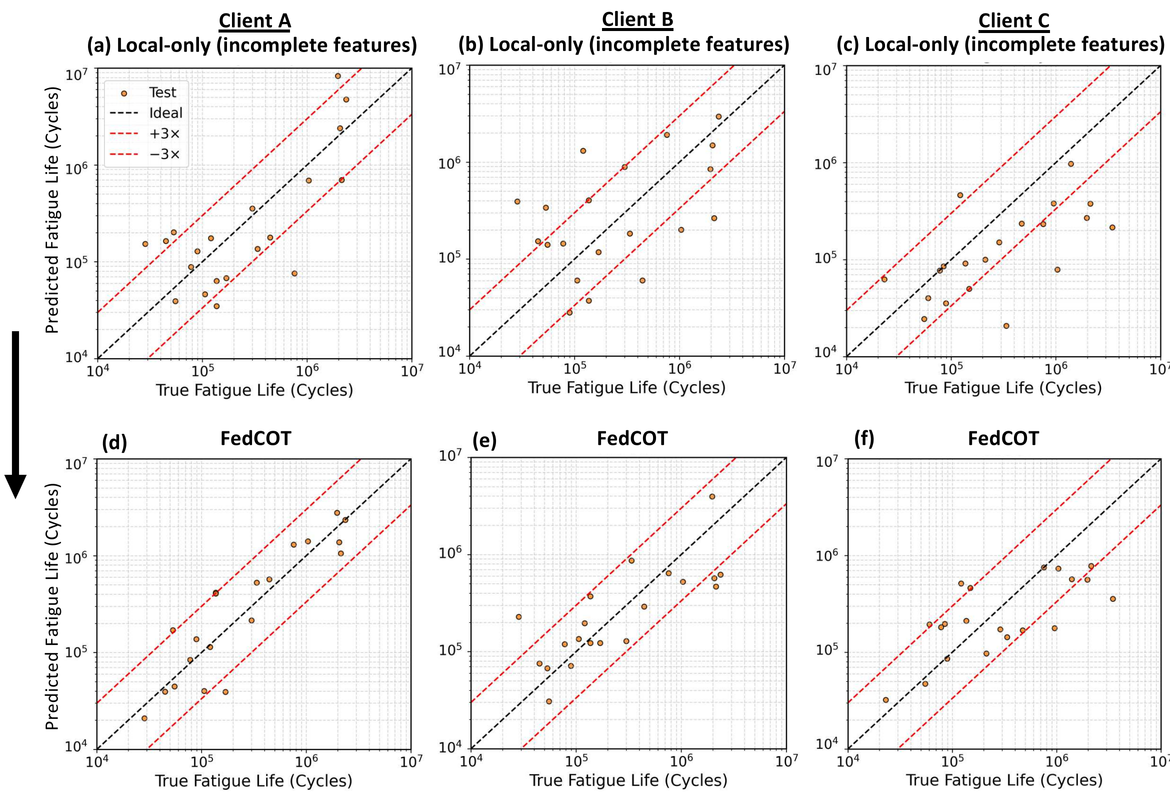

Figure 5.10: The prediction performance of predicted vs. true responses for (a)-(c) Local-only (incomplete features), and (d)-(f) FedCOT on test sets across different clients in the multi-client case study.

Table 5.8: Prediction performance comparison on the multi-client case study dataset. FedCOT consistently achieves the lowest error compared to Local-only (incomplete features), FedAvg, FedProx, Ditto, FedKL, and FedMMD. Reported values are mean $\pm$ std over 10 runs. ”IF” indicates ”Incomplete Features“. The value of ”-” means an invalid negative R^2.

Method	RMSE	MAPE	Adj. R^2	Residual Var.
Local-only (IF)	0.5852 (0.0934)	0.0987 (0.0282)	0.1452 (0.1170)	0.3275 (0.0553)
FedAvg	0.5272 (0.1061)	0.0784 (0.0144)	0.2426 (0.2067)	0.2671 (0.1050)
FedProx	0.4960 (0.0921)	0.0737 (0.0118)	0.2583 (0.2305)	0.2304 (0.0771)
Ditto	0.4648 (0.0939)	0.0680 (0.0122)	0.3937 (0.1827)	0.2023 (0.0898)
FedKL	0.6826 (0.1188)	0.1028 (0.0178)	-	0.4266 (0.1190)
FedMMD	0.6424 (0.0133)	0.0965 (0.0014)	-	0.3957 (0.0159)
FedCOT (Ours)	**0.3971 (0.0768)**	**0.0601 (0.0114)**	**0.5398 (0.1379)**	**0.1541 (0.0538)**
Gain (%)	↑ 34.57	↑ 21.70	↑ 271.76	↑ 38.11

with different levels of missing information. For Client A to C (severity of missing features increasing), the Local-only (IF) models exhibit large deviations from the ideal prediction line, especially for Client C. In comparison, FedCOT significantly improves predictions across all clients. These results indicate that FedCOT scales effectively to multi-client industrial settings and can leverage knowledge from the source client to improve fatigue life prediction performance for target clients without sharing sensitive data, even when critical defect-related features are unavailable.

5.5 Summary

This study introduces a personalized federated transfer learning framework with conditional optimal transport (FedCOT), allowing target clients with limited features to leverage knowledge of source clients with sufficient features without sharing private data. In FedCOT, each client trains an encoder-regressor structure through alternating optimization: the encoder projects heterogeneous inputs into a latent space and the regressor predicts responses based on these latent representations. Target clients align their latent representations with source clients', conditional on responses via COT, in a response-aware bin-wise manner. This work provided theoretical guarantees on OT gradients and predictive consistency. In addition, the simulations and a manufacturing case study showed that FedCOT achieved a significant improvement of 34.99% in RMSE compared to target clients that trained their models individually for fatigue life prediction of AM parts.

Chapter 6

Physics-informed Multiple Instance Learning for Multiaxial Fatigue Life Prediction in Additively Manufactured Metal Parts

Accurate prediction of multiaxial fatigue life is significant for the reliable application of laser powder bed fusion (L-PBF) components under realistic service loading conditions. Existing multiaxial fatigue models are mainly developed for conventional materials or assume that a single dominant critical defect governs fatigue failure. However, L-PBF components often contain multiple defects that can jointly contribute to fatigue failure. To address this challenge, this study proposes a Physics-informed Multiple Instance Learning (PhysMIL) framework for multiaxial fatigue life prediction in L-PBF parts. This methodology assumes each specimen as a bag of defects and integrates defect morphology, location, and multiaxial damage parameters within an attention-based learning framework. A physics-informed regularization derived from established fatigue mechanics is incorporated to lead model learning and improve interpretability. PhysMIL is validated through a numerical simulation and a public experimental study. Results demonstrate its effectiveness, achieving a 46.9% reduction of root mean square error compared with a baseline model. This framework provides a pathway that integrates fatigue mechanics and machine learning for reliable multiaxial fatigue assessment of L-PBF components.

6.1 Introduction

Accurately predicting the fatigue life of additive-manufactured (AM) metal parts under multiaxial loading is essential for the reliable adoption of parts in the aerospace, biomedical, and automotive industries [67]. Compared with uniaxial loading, multiaxial loading introduces complex stress states and varying crack propagation paths, making fatigue behavior significantly more difficult to predict. This challenge is pronounced in laser powder bed fusion (L-PBF) metals, where volumetric defects such as lack-of-fusion pores and keyholes act as local stress concentrators and dominate crack initiation [209].

Conventional multiaxial fatigue life prediction models, such as the mechanics-based models [67], critical plane approaches [40], or energy-based multiaxial theories [214], have been successfully explored in describing crack initiation in homogeneous, wrought alloys and applied to AM metals. While these models provide valuable physical insight, they rely on simplified defect representation or assume a dominant critical defect. In practice, however, AM parts often contain multiple defects, and fatigue failure may be affected by the combined effect of several defects rather than a single largest defect. As a result, traditional approaches may not fully capture the complex interaction between defects and multiaxial loading conditions.

Recent research on multiaxial fatigue life prediction using machine learning has made it possible to utilize a data-driven method for fatigue modeling using defect features extracted from XCT. Nevertheless, existing studies aggregate defect information using predefined statistical measures, such as average defect size. More importantly, fatigue life is measured at the specimen level, whereas defects exist at the defect level. This motivates us to model fatigue life prediction as a multiple instance learning (MIL) problem. Although MIL provides a mechanism for learning from collections of defect instances, existing MIL methods are mainly data-driven and lack physics knowledge [215, 216]. As a result, the learned relationships between fatigue life, loading conditions, and defect characteristics may violate established fatigue mechanisms, especially when training data are limited.

To address these challenges, this work proposes a Physics-informed Multiple Instance Learning (PhysMIL) framework for multiaxial fatigue life prediction of AM metal parts is proposed. PhysMIL treats each specimen as a bag and the defects within the specimen as instances. An attention mechanism is introduced to learn the defect importance and aggregate the defect-level information into a specimen-level representation. Moreover, a physics-informed regularization is employed to enforce a physically meaningful relationship between fatigue life, loading severity, and defect characteristics. This work evaluates PhysMIL using both numerical simulations and a multiaxial fatigue public dataset of L-PBF 316L stainless steel. Experimental results show that

PhysMIL predicts multiaxial fatigue life accurately while providing interpretable insights into defect criticality under different loading conditions.

6.2 Methodology

In this section, a PhysMIL framework for multiaxial fatigue life prediction of AM parts is proposed. The overall structure of the framework is illustrated in Figure 6.1. Given multiple defects observed from XCT inspection and specimen-level loading features, PhysMIL first encodes defect features into latent defect severity representations and then employs an attention mechanism to aggregate defect-level information. The aggregated defect representation is then combined with the multiaxial fatigue damage parameter and used for fatigue life prediction through a regression network. Additionally, physics-informed constraints are incorporated during training based on established relationships between loading conditions, defect characteristics, and fatigue life.

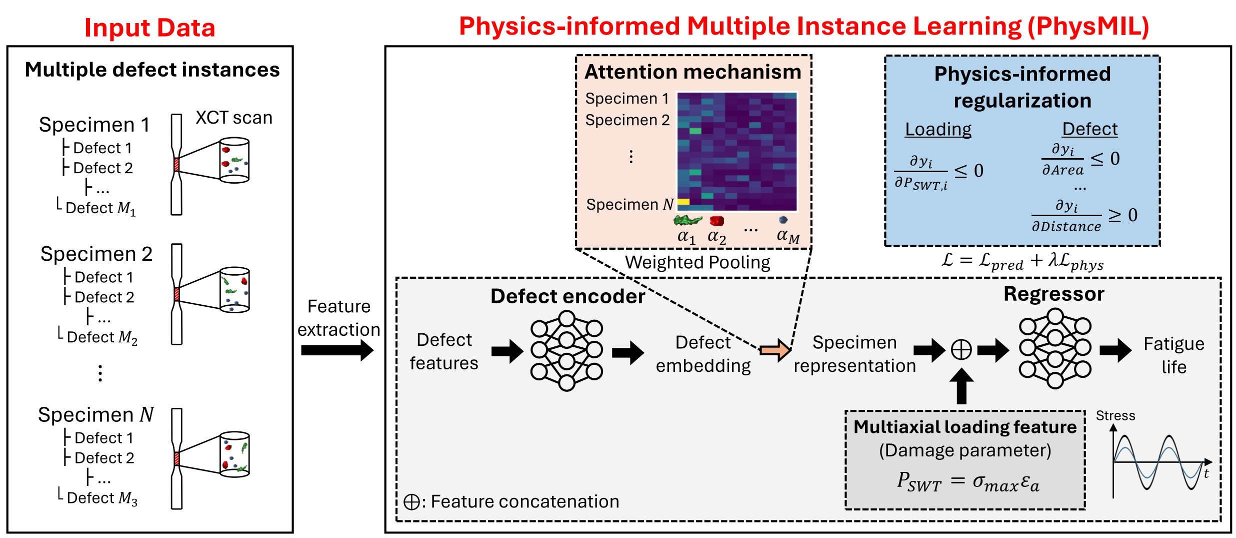

Figure 6.1: The framework of PhysMIL for multiaxial fatigue life prediction of L-PBF parts. For the input data, each specimen is treated as a bag with multiple defects. The extracted defect features and loading feature are processed by a defect encoder and regressor for fatigue life prediction. An attention mechanism is employed to learn defect importance and identify critical defects. Additionally, a physics-informed regularization term derived from established fatigue mechanics guides model learning and improves its interpretability.

6.2.1 Problem formulation

This work formulates the multiaxial fatigue life prediction of AM parts as a multiple-instance regression problem [217]. Specifically, each specimen contains multiple defects, and the fatigue life is only measured at the specimen level. Let $\mathcal{D} = \{(\mathcal{B}_i, \mathbf{s}_i, y_i)\}_{i=1}^{N}$ denote the dataset containing N specimens. For specimen i, $\mathcal{B}_i = \left\{\mathbf{x}_{ij}\right\}_{j=1}^{M_i}$ represents a bag with M_i defects, where $\mathbf{x}_{ij} \in \mathbb{R}^p$ is the extracted defect features. The specimen-level loading condition feature is represented by $\mathbf{s}_i \in \mathbb{R}^q$, and the corresponding fatigue life is $y_i \in \mathbb{R}$.

For the AM multiaxial fatigue dataset used in this study [84], the loading feature $\mathbf{s}_i = [P_{SWT,i}]$, where P_{SWT} is the Smith–Waston–Topper damage parameter, which is a widely used fatigue damage model that accounts for the detrimental effects of mean stress on the fatigue life of metals. It can effectively evaluate crack initiation under multiaxial loading conditions [84]. Each defect j is characterized by $\mathbf{x}_{ij} = [A_{ij}, \Psi_{ij}, L_{ij}, D_{ij}]$, where A_{ij}, Ψ_{ij}, L_{ij}, and D_{ij} denote projected defect area, sphericity, distance from the free surface, and equivalent defect diameter extracted from XCT scans, respectively. The objective is to learn the relationship $f_\Theta : (\mathcal{B}_i, \mathbf{s}_i) \to y_i$, where Θ is the regression model parameters. The learning problem is formulated as

$$\min_{\Theta} \frac{1}{N} \sum_{i=1}^{N} \ell\left(f_\Theta(\mathcal{B}_i, \mathbf{s}_i), y_i\right), \tag{6.1}$$

where $\ell(\cdot)$ is the loss function. This work adopts the mean square error (MSE) in this study.

The key challenge is that the critical defects are unknown, and multiple defects may jointly contribute to crack initiation and fatigue failure.

6.2.2 Physics-informed multiple instance learning

To address these challenges, this work proposes a PhysMIL framework. It can learn the relationship between defects and fatigue life from a set of defects while incorporating physics constraints, and provide potentially critical defects.

The overall architecture consists of three modules: (1) Defect encoding, (2) Attention-based defect aggregation and regressor, and (3) Physics-informed regularization.

Defect encoding

Each defect is encoded into a latent space to represent the defect severity by an encoder:

$$\mathbf{h}_{ij} = \phi(\mathbf{x}_{ij}) = \tanh\left(\mathbf{W}_e \mathbf{x}_{ij} + \mathbf{b}_e\right), \tag{6.2}$$

where $\mathbf{h}_{ij} \in \mathbb{R}^d$ denotes the latent representation of defect j in specimen i, and $\phi(\cdot)$ represents the encoder via a neural network.

Attention-based defect aggregation and regressor

This work utilizes an attention mechanism to learn the importance of defects since they contribute differently to fatigue failure. The attention score is represented as $r_{ij} = \mathbf{w}_a^T \mathbf{h}_{ij} + \mathbf{b}_a$. The corresponding attention weight is learned through a softmax function:

$$\alpha_{ij} = \frac{\exp(r_{ij})}{\sum_{k=1}^{M_i} \exp(r_{ik})}, \tag{6.3}$$

where

$$\alpha_{ij} \geq 0, \qquad \sum_{j=1}^{M_i} \alpha_{ij} = 1. \tag{6.4}$$

The specimen-level defect representation is

$$\mathbf{z}_i = \sum_{j=1}^{M_i} \alpha_{ij} \mathbf{h}_{ij}. \tag{6.5}$$

Then $\mathbf{z}_i$ and the loading feature were concatenated as $\mathbf{u}_i = [\mathbf{z}_i, \mathbf{s}_i]$. The fatigue life prediction is obtained by $\hat{y}_i = g_{\theta_g}(\mathbf{u}_i)$, where $g(\cdot)$ denotes the regressor for the prediction.

Importantly, traditional pooling operators such as mean pooling implicitly assume that all defects contribute equally to fatigue failure. Nevertheless, fatigue crack initiation is typically dominated by a subset of severe defects [15, 218]. Attention-based aggregation enables the model to learn defect importance weights directly from each specimen, thereby allowing adaptive identification of critical defects. The attention weights α_{ij} provide an interpretable measure of defect importance. Defects with higher attention weights contribute more significantly to the fatigue failure and might be treated as potential crack-initiation candidates.

Physics-informed regularization

This work incorporates physics-informed constraints to guide the learning of the proposed PhysMIL framework. Based on the established defect-fatigue knowledge [11, 105], fatigue life is expected to decrease with increasing loading severity $P_{SWT,i}$, projected defect area A_{ij}, and defect diameter L_{ij}, and decreasing defect sphericity Ψ_{ij} and distance from the free surface D_{ij}. The formulation of these relationships is shown as

$$\frac{\partial \hat{y}_i}{\partial P_{SWT,i}} \leq 0, \tag{6.6}$$

$$\frac{\partial \hat{y}_i}{\partial A_{ij}} \leq 0, \tag{6.7}$$

$$\frac{\partial \hat{y}_i}{\partial D_{ij}} \leq 0, \tag{6.8}$$

$$\frac{\partial \hat{y}_i}{\partial \Psi_{ij}} \geq 0, \tag{6.9}$$

$$\frac{\partial \hat{y}_i}{\partial L_{ij}} \geq 0. \tag{6.10}$$

To regulate and enforce these physical trends during training, a gradient-based regularization term in specimen and defect levels is developed:

$$\mathcal{L}_{phys} = \mathcal{L}_{spec} + \mathcal{L}_{def}, \tag{6.11}$$

where $\mathcal{L}_{spec}$ is the specimen-level physics loss in terms of the loading feature as

$$\mathcal{L}_{spec} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\partial \hat{y}_i}{\partial P_{SWT,i}}\right]_+^2, \tag{6.12}$$

and $\mathcal{L}_{def}$ is the defect-level physics loss regarding the defect features as

$$\mathcal{L}_{def} = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\left(\left[\frac{\partial \hat{y}_i}{\partial A_{ij}}\right]_+^2 + \left[-\frac{\partial \hat{y}_i}{\partial \Psi_{ij}}\right]_+^2 + \left[-\frac{\partial \hat{y}_i}{\partial L_{ij}}\right]_+^2 + \left[\frac{\partial \hat{y}_i}{\partial D_{ij}}\right]_+^2\right), \tag{6.13}$$

where $[x]_+ = \max(x, 0)$ is the rectified linear unit (ReLU) function that acts as the switch to control the penalty of each feature in the loss function.

Training objective and optimization

The proposed PhysMIL framework is trained in an end-to-end manner by jointly optimizing the defect encoder, attention aggregation module, and fatigue life regressor. Let $\Theta = \{\mathbf{W}_e, \mathbf{b}_e, \mathbf{w}_a, \mathbf{w}_a, \theta_g\}$ be all trainable model parameters, where $(\mathbf{W}_e, \mathbf{b}_e)$ are the encoder parameters, $(\mathbf{w}_a, \mathbf{b}_a)$ are the attention parameters, and θ_g denotes the regressor parameters.

To minimize the discrepancy between the predicted and observed fatigue life, the prediction loss is defined as the MSE

$$\mathcal{L}_{pred} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2, \tag{6.14}$$

where $\hat{y}_i$ and y_i denote the predicted and observed fatigue life of specimen i, respectively.

To enforce consistency with established fatigue mechanisms, the physics-informed regularization term $\mathcal{L}_{phys}$ is incorporated into the training objective. The overall loss function is defined as

$$\mathcal{L} = \mathcal{L}_{pred} + \lambda\mathcal{L}_{phys}, \tag{6.15}$$

where $\lambda > 0$ is a hyperparameter to control the trade-off between prediction accuracy and physical consistency.

The optimal model parameters are obtained by solving

$$\Theta^* = \arg\min_{\Theta} \mathcal{L}. \tag{6.16}$$

Since all components of PhysMIL are differentiable, gradients of the overall objective can be obtained through automatic differentiation and backpropagation throughout the network. The optimization is performed using the Adam optimizer [183]. At iteration t, the model parameters are updated as

$$\Theta^{(t)} = \Theta^{(t-1)} - \eta \, \mathrm{Adam}\left(\nabla_{\Theta}\mathcal{L}\right), \tag{6.17}$$

where η denotes the learning rate.

The proposed framework jointly learns defect representations, defect importance weights, and fatigue behavior while maintaining consistency with the known relationship between defects, loading, and fatigue life.

6.3 Simulation Study

To evaluate the performance of the proposed PhysMIL framework, a synthetic dataset for multiple instance regression that mimics the characteristics of the multiaxial fatigue behavior of AM parts was generated. The objective of this study is: (1) to investigate whether PhysMIL can accurately predict simulated multiaxial fatigue life from a set of defect instances, (2) to evaluate its capability to identify critical defects, and (3) to quantify the importance of the proposed physics-informed regularization.

The proposed PhysMIL framework was compared with three representative baseline methods:

- **Global-only Model**: A traditional physics-based prediction using only the simulated loading feature (P_{SWT}) without defect information.

- **Physics-informed Neural Network (PINN) [84]**: A physics-informed neural network that utilizes both loading and defect features. Multiple defect information is aggregated by a convolutional neural network, and then integrated with the fatigue damage parameter for prediction.

- **AttentionMIL**: Standard attention-based multiple instance learning without physics-informed regularization.

This work investigates the model performance of all benchmark models via R^2, root mean square error (RMSE), and symmetric mean absolute percentage error (SMAPE). SMAPE can avoid the asymmetric bias when the true values are small. Moreover, critical defect identification performance was assessed by comparing the designed attention and learned attention from PhysMIL. All models are validated via 5-fold cross-validation.

In this synthetic multiaxial fatigue dataset, each specimen (i.e., bag) contains one loading feature (i.e., global feature) and multiple defects (i.e., instances) described by defect features. The fatigue life (response) is generated according to the hidden defect severity mechanism, where both loading conditions and defect characteristics contribute to the fatigue life.

For each specimen, they are represented as a bag containing multiple defect instances together with a specimen-level loading feature. The generated data satisfy three key properties: (1) the response is only observed at the specimen level, (2) different defect instances contribute unequally to the response, and (3) physical relationships exist between input variables and the response.

A total of $N = 100$ specimens were generated, with each specimen containing $M = 10$ defect instances. Each defect has three features, which are represented by a feature vector $\mathbf{x}_{ij} \in \mathbb{R}^3$. The features are sampled from a multivariate Gaussian distribution $\mathbf{x}_{ij} \sim \mathcal{N}(\mu, \Sigma)$ with mean $\boldsymbol{\mu} = [0, 0, 0]^T$ and standard deviation

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & 0.3 & 0.2 \\ 0.3 & 1 & 0.4 \\ 0.2 & 0.4 & 1 \end{bmatrix}. \tag{6.18}$$

The correlated covariance structure introduces statistical dependence among defect features, which mimics realistic AM defect data.

Additionally, each specimen contains a loading feature $\mathbf{s}_i \sim \text{LogNormal}(\mu_S, \sigma_S^2)$ with $\mu_S = 0.35$ and $\sigma_S = 0.6$, which aligns with the fitted distribution from real-world damage parameter P_{SWT} data. For each defect instance, a latent defect risk r_{ij} is defined as the softplus function $\log\left(1 + \exp(\mathbf{w}^T \mathbf{x}_{ij} + b)\right)$, where $w = [1.0, -1.2, 0.9]^T$ and $b = 0.1$. It can provide a non-negative defect severity measure and preserve monotonic relationships. The latent risk is the contribution of an individual defect to specimen failure. The coefficients establish monotonic relationships between the defect features and the latent risk, thereby developing physically interpretable instance-level importance. The ground-truth attention distribution can be determined as $r_{ij}^{\text{true}} = \frac{\exp(r_{ij})}{\sum_{k=1}^{M} \exp(r_{ik})}, \sum_{j=1}^{M} r_{ij}^{\text{true}} = 1$. This ground-truth attention distribution enables direct evaluation of whether the learned attention mechanism can recover defect importance. The latent risks are then aggregated into a specimen-level risk through the combination of dominant defect and multi-defect accumulation mechanisms:

$$R_i = \eta \max_{1 \le j \le M} r_{ij} + (1-\eta) \log\left(\frac{1}{M_i} \sum_{j=1}^{M_i} \exp(r_{ij})\right), \qquad 0 \le \eta \le 1, \tag{6.19}$$

where R_i denotes the specimen-level defect risk and η controls the relative contribution of the two aggregation mechanisms. The first term represents a dominant-defect mechanism, in which the specimen response is governed primarily by the single critical defect. This assumption is commonly adopted in defect-based fatigue assessment, where crack initiation is often associated with the most severe defect within a specimen. In contrast, the second term corresponds to a multi-defect accumulation mechanism, where multiple defects jointly contribute to specimen degradation. The log-sum-exp formulation provides a smooth approximation that accounts for the influence of all defects while assigning greater weight to defects with higher risk.

By combining the two terms, the simulation can represent a broad range of failure behaviors, ranging from single-defect-dominated failure ($\eta = 1$) to multi-defect accumulation ($\eta = 0$). In this study, $\eta = 0.5$ is used to create a mixed regime in which both mechanisms coexist. This setting provides a challenging scenario for evaluating whether the proposed PhysMIL framework can simultaneously identify critical defects and capture the influence of defect populations.

Finally, the response is generated as

$$y_i = \alpha_0 - \alpha_{\mathbf{s}} \log(s_i) - \alpha_R R_i + \epsilon_i \tag{6.20}$$

where s_i denotes the specimen-level loading feature, R_i is the aggregated defect risk, and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ is the additive Gaussian noise. In this simulation, this work sets $a_0 = 5.0$, $a_s = 1.2$, $a_R = 1.0$, and $\sigma = 0.2$.

This formulation ensures that the response decreases monotonically with increasing loading severity and defect risk, thereby preserving the physically meaningful relationships embedded in the simulation. The additive noise term introduces stochastic variability and measurement uncertainty, providing a more realistic regression task. Consequently, the generated dataset enables systematic evaluation of prediction accuracy, defect-importance recovery, and the effectiveness of physics-informed regularization.

6.3.1 Simulation results

Table 6.1 shows the prediction performance of all models, and Figure 6.2 demonstrates the corresponding prediction results. It is noted that PhysMIL outperforms Global-only, PINN, and AttentionMIL, with an average R^2 of 0.904, RMSE of 0.255, and SMAPE of 9.097%. Compared with the Global-only model, it improves the prediction accuracy by 0.506 in R^2 while reducing the RMSE and SMAPE by 60.0% and 55.8%, respectively. Furthermore, compared with AttentionMIL, the additional improvement demonstrates that incorporating physics-informed constraints can further guide the learning process toward physically meaningful representations.

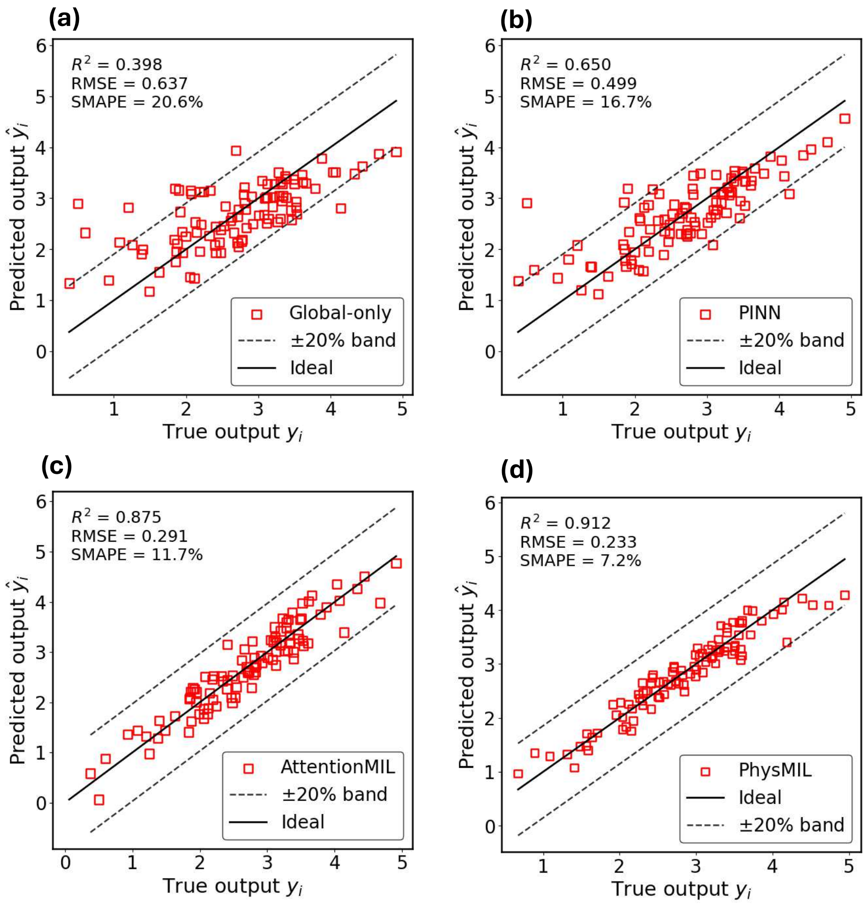

Figure 6.2: Comparison of prediction performance for the simulated dataset: (a) Global-only Model, (b) PINN, (c) AttentionMIL, and (d) PhysMIL. The dashed lines denote the ±20% error bounds, and the solid line represents the ideal prediction.

To investigate the defect instance importance, Figure 6.3 shows the comparison between the ground-truth attention map and the attention weights learned by PhysMIL. It can be observed that the learned attention map closely matches the underlying defect importance pattern. In particular, the dominant defects receiving the highest ground-truth attention are consistently assigned the largest learned attention weights. These findings validate the effectiveness of combining attention-based mechanisms with physics-informed regularization for simulated defect-driven fatigue modeling.

Table 6.1: Performance Comparison of PhysMIL and Baseline Models on the Simulated Dataset.

Model	R^2	R^2 **gain**	**RMSE**	**RMSE reduction**	**SMAPE (%)**	**SMAPE reduction**
Global-only	0.398 ± 0.200	—	0.637 ± 0.091	—	20.598 ± 4.496	—
PINN	0.650 ± 0.083	$+0.252$	0.499 ± 0.122	21.7%	16.747 ± 4.712	18.7%
Attention-MIL	0.875 ± 0.039	$+0.477$	0.291 ± 0.041	54.3%	11.691 ± 4.202	43.2%
PhysMIL	0.912 ± 0.019	$+0.514$	0.233 ± 0.046	**63.42%**	7.167 ± 1.232	**65.21%**

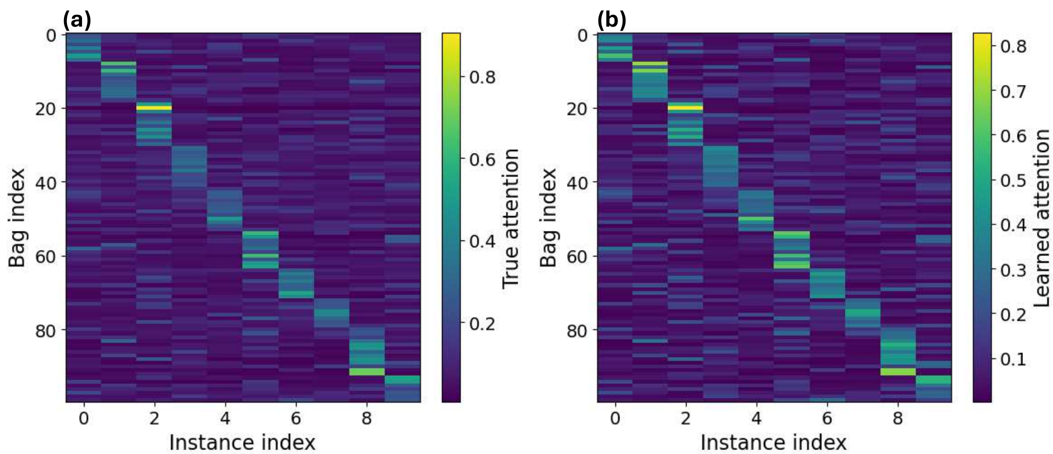

Figure 6.3: Attention map comparison between (a) ground-truth attention map and (b) learned attention map. It shows that the learned map closely matches the underlying defect importance.

6.3.2 Ablation and sensitivity analysis

This work conducted an ablation study to quantify the individual component contributions in the proposed PhysMIL framework and performed sensitivity analysis with respect to the sample size.

Ablation study of PhysMIL components

This study investigates the impact of incorporating the global feature, attention mechanism, and physics constraints. The same simulated dataset and training settings described in Section 6.3.1 were used for all model variants.

In Table 6.2, PhysMIL achieves the best performance, which incorporates global features, an attention mechanism, and physics constraints. Compared with AttentionMIL, PhysMIL improves the average R^2 from 0.875 to 0.904 while reducing the RMSE SMAPE by around 12.4% and 22.2%, respectively. These results indicate that the physics-informed constraints provide complementary information beyond attention-based defect instance aggregation and offer additional physical guidance during model training.

Overall, the ablation study confirms that each component is important to the final prediction performance. The global features capture specimen-level fatigue severity, the attention mechanism estimates the defect importance, and the physics-informed regularization further guides the model training.

Table 6.2: Ablation Study of PhysMIL Components on the Simulated Dataset. This evaluates the contribution of each key component of the PhysMIL framework: global feature fusion, attention mechanism, and physics constraints.

Ablation Settings				**Model Performance**		
Configuration	Global Feature	Attention Mechanism	Physics Constraint	R^2 ↑	RMSE ↓	SMAPE (%) ↓
Mean Pooling	×	×	×	—	1.274 ± 0.174	37.315 ± 2.890
Mean Pooling + Global	✓	×	×	0.825 ± 0.047	0.356 ± 0.108	14.203 ± 6.163
Attention-MIL	✓	✓	×	0.875 ± 0.039	0.291 ± 0.041	11.691 ± 4.202
PhysMIL	✓	✓	✓	0.912 ± 0.019	0.233 ± 0.046	7.167 ± 1.232

Sensitivity to sample size

Fatigue samples in the real-world AM datasets are usually limited due to the high cost of specimen fabrication and time-consuming fatigue testing. To investigate the robustness of PhysMIL under different data availability conditions, a sensitivity analysis was performed using four sample sizes ($N = 30, 50, 100,$ and 200).

Table 6.3: Robustness of PhysMIL under different training sample sizes on the simulated dataset.

Sample Size N	$R^2 \uparrow$	RMSE $\downarrow$	SMAPE (%) $\downarrow$
30	0.711 ± 0.199	0.437 ± 0.151	16.820 ± 5.728
50	0.883 ± 0.039	0.310 ± 0.071	12.841 ± 7.120
100	0.912 ± 0.019	0.233 ± 0.046	7.167 ± 1.232
200	0.936 ± 0.017	0.245 ± 0.018	10.325 ± 2.002

Table 6.3 shows that with the increase of the sample size, the model performance gradually improves. The average R^2 increases from 0.711 at $N = 30$ to 0.936 at $N = 200$, while RMSE decreases from 0.437 to 0.245. More importantly, PhysMIL maintains a relatively reasonable predictive capability even under limited-data conditions. With only 30 samples, PhysMIL still performs well, indicating its ability to learn meaningful relationships between simulated loading features, defect characteristics, and the response. The results claim that PhysMIL exhibits good scalability in terms of sample size.

6.4 Case Study of Multiaxial Fatigue Life Prediction in AM Parts

6.4.1 Problem Description and Data Setup

To assess the effectiveness of the proposed PhysMIL framework on a real-world multiaxial fatigue life prediction in AM, a case study was conducted via a public multiaxial fatigue dataset of L-PBF manufactured 316L stainless steel reported in [84].

The dataset consists of 20 thin-walled tubular specimens under strain-controlled multiaxial low-cycle fatigue loading under axial, torsional, and proportional loading paths, as shown in Figure 6.4. For each specimen, fatigue life was measured as the number of cycles to failure, while XCT was used to characterize internal volumetric defects prior to fatigue testing.

This dataset provides an ideal benchmark for evaluating multiple instance learning approaches since each specimen has multiple candidate defects that might contribute to fatigue failure. Each specimen contains 10 selected defects with the largest projected area for analysis. Defects are quantified by 4 features: projected defect area, equivalent defect diameter, sphericity, and distance to the free surface. In addition to defect-level information, the specimen-level multiaxial loading condition was represented by the Smith-Watson-Topper (SWT) damage parameter $P_{SWT} = \sigma_{\max}\varepsilon_a$, where $\sigma_{\max}$ and ε_a are the maximum normal stress and normal strain amplitude acting on the critical plane. It has been widely used for low-cycle multiaxial fatigue assessment and serves as the global loading feature in PhysMIL.

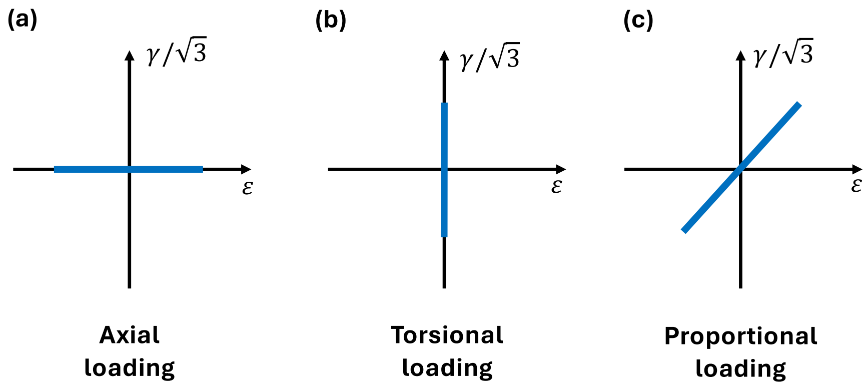

Figure 6.4: Representative multiaxial loading paths considered in the fatigue dataset [84]: (a) axial loading, (b) torsional loading, and (c) proportional axial–torsional loading.

This work adopts leave-one-out cross-validation (LOOCV) to maximize data utilization while offering an unbiased estimate of predictive performance on this small-sample dataset. At each iteration, one specimen was used for testing, and the remaining specimens were used for model training. The proposed PhysMIL was compared with three models, P_{SWT}-only, PINN, and AttentionMIL. The prediction performance was evaluated using R^2, RMSE, and SMPAE, the same as the simulation.

6.4.2 Results

Multiaxial fatigue life prediction results

Table 6.4 and Figure 6.5 show the comparison of the multiaxial fatigue life prediction performance for all models via LOOCV. It is noted that PhysMIL achieves the best predictive performance among all benchmark models, indicating the effectiveness of incorporating defect information and physics-informed regularization. It got an R^2 of 0.856, an RMSE of 0.221, and an SMAPE of 7.0%. Compared with P_{SWT}-only model, PhysMIL improves the R^2 by 0.366 and reduces the RMSE by 46.9%. Compared with AttentionMIL, the RMSE in PhysMIL is further reduced by 15.3%, demonstrating the benefit of incorporating physics-informed regularization into the multiple instance learning framework. Additionally, predictions in PhysMIL are more tightly clustered around the ideal prediction line, indicating the strong robustness even with limited specimens.

Table 6.4: LOOCV prediction performance comparison on the public multiaxial fatigue life dataset.

Model	R^2	R^2 **gain**	**RMSE**	**RMSE reduction**	**SMAPE (%)**	**SMAPE reduction**
P_{SWT}-only	0.490	—	0.416	—	15.800	—
PINN	0.695	+0.205	0.322	22.6%	12.100	23.4%
Attention-MIL	0.799	+0.309	0.261	37.0%	10.300	34.8%
PhysMIL	**0.856**	**+0.366**	**0.221**	**46.9%**	**7.000**	**55.7%**

Defect attention analysis

Furthermore, this work evaluates the learned defect importance from the attention map. Figure 6.6 visualizes the attention weights assigned to the ten defects within each specimen by the proposed PhysMIL framework. First, the learned attention distributions are highly non-uniform, which shows that the model focuses on defects that are most detrimental to fatigue failure. Second, PhysMIL dynamically determines defect importance based on the combined effect of defect size, morphology, location, and loading conditions. This behavior is consistent with experimental observations that fatigue crack initiation is influenced by multiple defect characteristics rather than defect size alone. Third, many specimens exhibit distributed attention across multiple defects instead of focusing all attention on a dominant critical defect. This finding supports that a single critical defect might not always govern fatigue behavior. Rather, multiple defects can jointly contribute to local stress concentration and crack initiation.

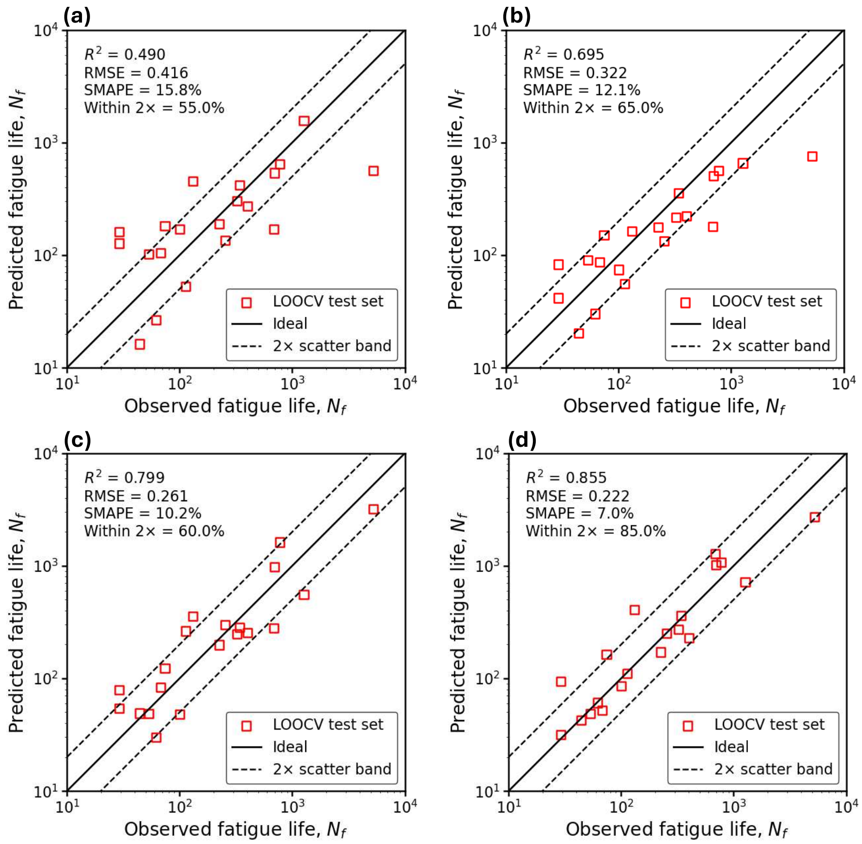

Figure 6.5: LOOCV fatigue life prediction results for (a) P_{SWT}-only model, (b) PINN, (c) Attention-MIL, and (d) PhysMIL. The solid line denotes the ideal prediction, and the dashed lines indicate the factor-of-two scatter band.

Physics consistency analysis

Then, this work assesses whether the proposed physics-informed regularization successfully guides the model learning. The monotonicity violation rates of the learned model were quantified for each input feature, as shown in Figure 6.7. The P_{SWT} damage parameter exhibits a violation rate of 0%, indicating that the learned relationship between loading severity and fatigue life is consistent with the physics that increasing loading severity reduces fatigue life. Among the defect features, the projected area shows the lowest violation rate of 2.1%, suggesting that the model learns the detrimental influence of larger defects. In contrast, distance from the free surface and sphericity show higher violation rates of 69.0% and 76.0%, respectively. In practice, defect morphology and spatial location interact with other factors such as loading conditions

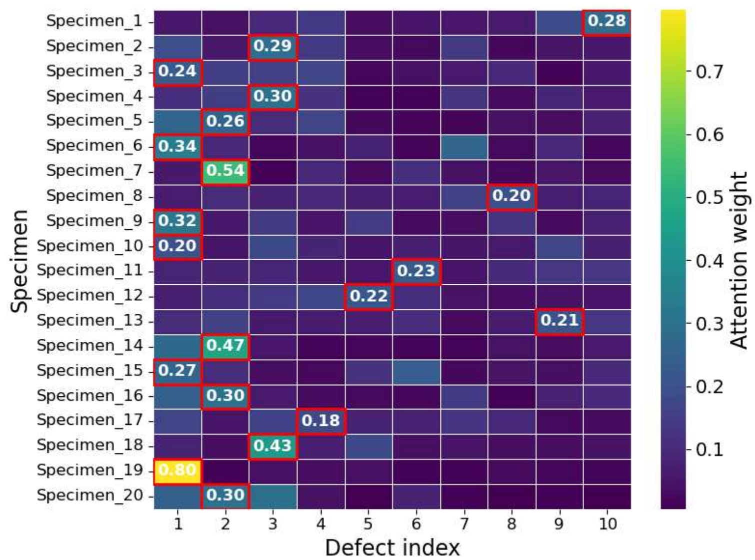

Figure 6.6: Attention weight distributions learned by PhysMIL for the ten defects within each specimen. Higher attention weights indicate defects that contribute more strongly to the predicted fatigue life. Note: for each specimen, defects are ranked based on projected area from largest to smallest. Defect Index 1 denotes the largest defect, whereas Defect Index 10 is the smallest defect.

and neighboring defects, making their effects more difficult to learn from data alone.

Loading-path-dependent attention behavior

Figure 6.8 shows how the learned attention distribution changes under different loading paths by evaluating the attention entropy of each specimen. Lower entropy means that the model concentrates attention on a particular critical defect, and higher entropy indicates that it focuses on multiple defect mechanisms. As shown in Figure 6.8 (a) and (b), attention entropy shows significant variability across specimens under different loading conditions. Torsional loading displays the largest variability in attention entropy, but proportional loading and axial loading exhibit relatively low attention entropy.

These observations provide two key pieces of information: (1) fatigue behavior under axial loading and proportional loading may involve multiple defect contributions, not only one dominant critical defect, (2) fatigue behavior under torsional loading may sometimes be strongly single-defect dominated, depending on the defect population. They are consistent

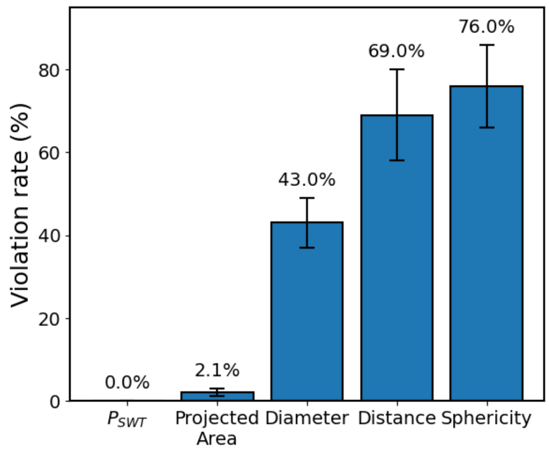

Figure 6.7: Monotonicity violation rates of the learned PhysMIL model regarding the loading feature (P_{SWT}) and defect features. Lower violation rates indicate better agreement with the imposed physics-informed regularization.

with the underlying fatigue physics. Under axial and proportional loading, the principal stress direction is relatively stable throughout the loading cycle, allowing multiple defects to experience comparable stress concentration effects and contribute to crack initiation. In contrast, torsional loading is governed primarily by shear stresses, and defect criticality becomes highly dependent on defect geometry and location relative to the local shear field. As a result, certain defects might become significantly more critical than others, causing the model to concentrate attention on a small subset of defects. Overall, these findings indicate that the attention mechanism captures physically meaningful differences in defect importance under different loading paths and provides additional insight into the defect-driven fatigue behavior of AM materials.

6.5 Summary

This study proposes a PhysMIL framework for multiaxial fatigue life prediction of AM metal parts. The defect-driven fatigue life prediction is formulated as a multiple instance learning problem, where each specimen is represented as a bag of defects and learns the relationship between fatigue life, loading severity, and defect characteristics from specimen-level and defect-level information. By integrating attention-based defect aggregation with physics-informed regularization, PhysMIL simultaneously learns defect importance, predicts fatigue life, and preserves

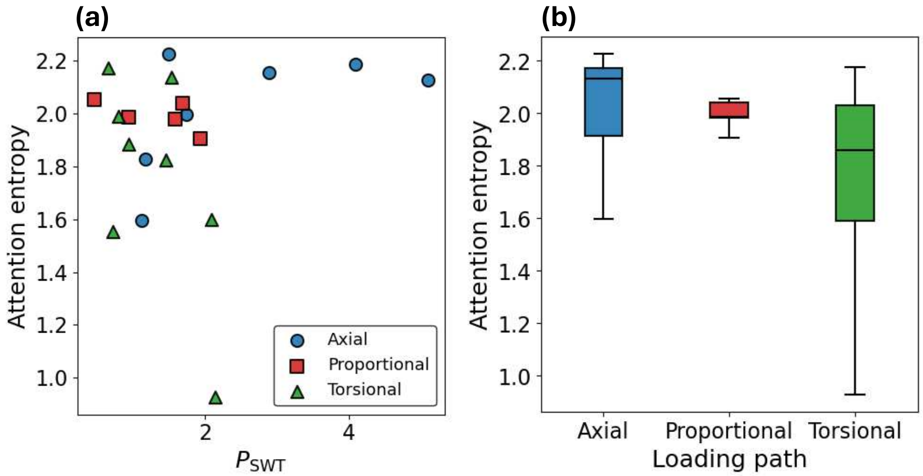

Figure 6.8: Analysis of attention entropy under different loading paths. (a) The scatter plot of attention entropy with respect to fatigue damage parameter P_{SWT}. (b) The boxplot of attention entropy for axial, proportional, and torsional loading conditions.

physically meaningful relationships between fatigue life, loading severity, and defect characteristics. The proposed framework was evaluated by both simulations and a public multiaxial fatigue dataset of L-PBF 316L stainless steel. The case study results showed that PhysMIL achieved a significant improvement of 0.366 in R^2 and 46.9% in RMSE compared to the traditional physics model. Beyond predictive performance, PhysMIL provides a new perspective for understanding defect-driven fatigue behavior in AM parts. The attention analysis revealed that fatigue failure may be governed by either a dominant defect or the combined contribution of multiple defects, depending on the loading condition and defect population.

Chapter 7

Conclusions and Future Work

7.1 Summary of Contributions

This research developed a physics-informed hierarchical learning framework for defect characterization and reliable fatigue life prediction in metal AM. The framework advances fatigue assessment across various hierarchical levels, evolving from defect understanding (defect-level), fatigue life prediction (specimen-level), privacy-preserving collaborative learning (organization level), to physics-informed predictive modeling under increasingly realistic loading conditions (multiple-defect level). The research addresses four fundamental challenges: (1) characterizing volumetric defects, (2) enabling nondestructive fatigue life prediction, (3) developing collaborative learning across organizations, and (4) generalizing fatigue life prediction under multiaxial loading conditions.

The first study developed an Integrated Data-driven Analytical Framework for Defect Criticality (IDADC) to quantify the impact of critical defect characteristics on fatigue life of L-PBF parts and proposed a novel critical directional curvature feature of defects from XCT. By integrating defect feature extraction, correlation analysis, machine learning, and model interpretation techniques, this work highlights the most important defect features, such as size, morphology, and location, as dominant factors governing fatigue failure, with a relatively strong correlation of over 0.5.

Building upon this understanding, the second study proposed a Multimodal Transfer Learning (MMTL) framework for nondestructive fatigue life prediction by linking process parameters, XCT-derived defect information, and fatigue life through a hierarchical graph convolutional network (HGCN) and knowledge transfer strategy. MMTL successfully transferred learned defect knowledge from data-rich XCT defects to data-limited fatigue datasets via a classification task, achieving an F1-score of 0.9593. It achieves an RMSLE of 0.3049, representing a 19.3% improvement compared to the baseline model.

The third study developed a Personalized Federated Transfer Learning Framework via Conditional Optimal Transport (FedCOT) for collaborative predictive modeling across organizations with an application on fatigue life prediction. FedCOT leverages COT to align heterogeneous feature spaces across clients, enabling organizations with different inspection capabilities to jointly improve prediction performance without sharing raw data. FedCOT demonstrated substantial improvement of 34.99% in fatigue life prediction of L-PBF parts over isolated local models, highlighting its potential to enable secure and scalable collaborative learning among manufacturers with differing inspection capabilities.

The fourth study further advances fatigue life prediction from conventional uniaxial loading to realistic multiaxial loading conditions via a Physics-informed Multiple Instance Learning (PhysMIL) framework. The proposed methodology models defect populations within each specimen and integrates physics knowledge with attention-based mechanics. This model not only achieves the lowest SMAPE of 7% on the case study, but also provides interpretable defect importance to reveal the defect-fatigue failure mechanism under different loading conditions. It establishes a foundation for generalized multiaxial fatigue life prediction in AM components.

This research demonstrates how physics-informed hierarchical learning can bridge the gap between defect characterization and deployable fatigue life prediction in AM.

7.2 Future Work

Building upon these studies, the proposed subsequent research focuses on several research opportunities:

First, uncertainty quantification of fatigue life prediction in AM will be investigated. Fatigue scatter is a common issue due to variability arising from manufacturing processes, geometry design, material properties, defect populations, and loading conditions. Integrating probabilistic modeling with physics-informed learning could provide confidence bounds for fatigue life predictions and improve decision-making in safety-critical applications.

Second, the current framework has the potential to be extended to cross-machine, cross-process, and cross-material generalization. Although this dissertation focuses on L-PBF metal parts, future studies may explore transfer learning strategies across different alloys, AM processes, and sensing modalities to establish more generalized fatigue models.

Finally, incorporating active learning and adaptive testing strategies in future work. By identifying the most important and informative parts for experimentation, machine learning models can guide data collection and reduce cost with respect to fatigue testing and XCT inspection.

[1] M. Soori, F. K. G. Jough, R. Dastres, and B. Arezoo, "Additive manufacturing modification by artificial intelligence, machine learning, and deep learning: a review," Additive Manufacturing Frontiers, 200198 (2025).

[2] M. Taghian, M. H. Mosallanejad, E. Lannunziata, G. Del Greco, L. Iuliano, and A. Saboori, "Laser powder bed fusion of metallic components: latest progress in productivity, quality, and cost perspectives," Journal of Materials Research and Technology **27**, 6484–6500 (2023).

[3] S. Mittal, M. A. Khan, D. Romero, and T. Wuest, "A critical review of smart manufacturing & industry 4.0 maturity models: implications for small and medium-sized enterprises (smes)," Journal of manufacturing systems **49**, 194–214 (2018).

[4] L. E. Criales, Y. M. Arısoy, B. Lane, S. Moylan, A. Donmez, and T. Özel, "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 (2017).

[5] H. Gong, K. Rafi, H. Gu, T. Starr, and B. Stucker, "Analysis of defect generation in ti–6al–4v parts made using powder bed fusion additive manufacturing processes," Additive Manufacturing **1**, 87–98 (2014).

[6] J. V. Gordon, S. P. Narra, R. W. Cunningham, H. Liu, H. Chen, R. M. Suter, J. L. Beuth, and A. D. Rollett, "Defect structure process maps for laser powder bed fusion additive manufacturing," Additive Manufacturing **36**, 101552 (2020).

[7] S. Pal, G. Lojen, R. Hudak, V. Rajtukova, T. Brajlih, V. Kokol, and I. Drstvenšek, "As-fabricated surface morphologies of ti-6al-4v samples fabricated by different laser processing parameters in selective laser melting," Additive Manufacturing **33**, 101147 (2020).

[8] H. Choo, K.-L. Sham, J. Bohling, A. Ngo, X. Xiao, Y. Ren, P. J. Depond, M. J. Matthews, and E. Garlea, "Effect of laser power on defect, texture, and microstructure of a laser powder bed fusion processed 316l stainless steel," Materials & Design **164**, 107534 (2019).

[9] W. M. Tucho, V. H. Lysne, H. Austbø, A. Sjolyst-Kverneland, and V. Hansen, "Investigation of effects of process parameters on microstructure and hardness of slm manufactured ss316l," Journal of Alloys and Compounds **740**, 910–925 (2018).

[10] A. Poudel, M. S. Yasin, J. Ye, J. Liu, A. Vinel, S. Shao, and N. Shamsaei, "Feature-based volumetric defect classification in metal additive manufacturing," Nature Communications **13**, 6369 (2022).

[11] A. Li, S. Baig, J. Liu, S. Shao, and N. Shamsaei, “Defect criticality analysis on fatigue life of l-pbf 17-4 ph stainless steel via machine learning,” International Journal of Fatigue **163**, 107018 (2022).

[12] A. Du Plessis, I. Yadroitsava, and I. Yadroitsev, “Effects of defects on mechanical properties in metal additive manufacturing: a review focusing on x-ray tomography insights,” Materials & Design **187**, 108385 (2020).

[13] M. Tang and P. C. Pistorius, “Fatigue life prediction for alsi10mg components produced by selective laser melting,” International journal of fatigue **125**, 479–490 (2019).

[14] N. Sanaei, A. Fatemi, and N. Phan, “Defect characteristics and analysis of their variability in metal l-pbf additive manufacturing,” Materials & Design **182**, 108091 (2019).

[15] S. Tammas-Williams, P. J. Withers, I. Todd, and P. Prangnell, “The influence of porosity on fatigue crack initiation in additively manufactured titanium components,” Scientific reports **7**, 7308 (2017).

[16] I. Maskery, N. Aboulkhair, M. Corfield, C. Tuck, A. Clare, R. K. Leach, R. D. Wildman, I. Ashcroft, and R. J. Hague, “Quantification and characterisation of porosity in selectively laser melted al–si10–mg using x-ray computed tomography,” Materials Characterization **111**, 193–204 (2016).

[17] G. Kasperovich, J. Haubrich, J. Gussone, and G. Requena, “Correlation between porosity and processing parameters in tial6v4 produced by selective laser melting,” Materials & Design **105**, 160–170 (2016).

[18] J. Ye, A. Poudel, J. Liu, A. Vinel, D. Silva, S. Shao, and N. Shamsaei, “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**, 3093–3107 (2023).

[19] R. Snell, S. Tammas-Williams, L. Chechik, A. Lyle, E. Hernández-Nava, C. Boig, G. Panoutsos, and I. Todd, “Methods for rapid pore classification in metal additive manufacturing,” Jom **72**, 101–109 (2020).

[20] S. Romano, A. Brückner-Foit, A. Brandão, J. Gumpinger, T. Ghidini, and S. Beretta, “Fatigue properties of alsi10mg obtained by additive manufacturing: defect-based modelling and prediction of fatigue strength,” Engineering Fracture Mechanics **187**, 165–189 (2018).

[21] A. Fatemi, H. O. Fuchs, R. I. Stephens, and R. R. Stephens, “Metal fatigue in engineering,” A wiley-interscience Publication (2001).

[22] J. Tavernelli and L. Coffin, “Experimental support for generalized equation predicting low cycle fatigue,” Journal of Basic Engineering **84**, 533–537 (1962).

[23] R. Wei, “Fracture mechanics approach to fatigue analysis in design,” (1978).

[24] S. Bhat and R. Patibandla, "Metal fatigue and basic theoretical models: a review," Alloy steel-properties and use **22**, 203–238 (2011).

[25] U. Zerbst, M. Madia, C. Klinger, D. Bettge, and Y. Murakami, "Defects as a root cause of fatigue failure of metallic components. i: basic aspects," Engineering Failure Analysis **97**, 777–792 (2019).

[26] H. Kitagawa, "Applicability of fracture mechanics to very small cracks or the cracks in the early stage," in Proc. 2nd int. conf. on mechanical behaviour of materials (1976), pp. 627–631.

[27] Y. Murakami, *Metal fatigue: effects of small defects and nonmetallic inclusions* (Academic Press, 2019).

[28] Y. Murakami and S. Beretta, "Small defects and inhomogeneities in fatigue strength: experiments, models and statistical implications," Extremes **2**, 123–147 (1999).

[29] F. Stern, J. Tenkamp, and F. Walther, "Non-destructive characterization of process-induced defects and their effect on the fatigue behavior of austenitic steel 316l made by laser-powder bed fusion," Progress in Additive Manufacturing **5**, 287–294 (2020).

[30] S. Romano, A. Brandão, J. Gumpinger, M. Gschweitl, and S. Beretta, "Qualification of am parts: extreme value statistics applied to tomographic measurements," Materials & Design **131**, 32–48 (2017).

[31] Y. Hu, S. Wu, Z. Wu, X. Zhong, S. Ahmed, S. Karabal, X. Xiao, H. Zhang, and P. Withers, "A new approach to correlate the defect population with the fatigue life of selective laser melted ti-6al-4v alloy," International Journal of Fatigue **136**, 105584 (2020).

[32] J.-R. Poulin, A. Kreitcberg, P. Terriault, and V. Brailovski, "Fatigue strength prediction of laser powder bed fusion processed inconel 625 specimens with intentionally-seeded porosity: feasibility study," International Journal of Fatigue **132**, 105394 (2020).

[33] B. Torries, A. Imandoust, S. Beretta, S. Shao, and N. Shamsaei, "Overview on microstructure- and defect-sensitive fatigue modeling of additively manufactured materials," Jom **70**, 1853–1862 (2018).

[34] D. Kotzem, S. Kleszczynski, F. Stern, A. Elspaß, J. Tenkamp, G. Witt, and F. Walther, "Impact of single structural voids on fatigue properties of aisi 316l manufactured by laser powder bed fusion," International Journal of Fatigue **148**, 106207 (2021).

[35] D.-F. Li, R. A. Barrett, P. E. O'Donoghue, N. P. O'Dowd, and S. B. Leen, "A multi-scale crystal plasticity model for cyclic plasticity and low-cycle fatigue in a precipitate-strengthened steel at elevated temperature," Journal of the Mechanics and Physics of Solids **101**, 44–62 (2017).

[36] B. Smith, D. Shih, and D. McDowell, “Cyclic plasticity experiments and polycrystal plasticity modeling of three distinct ti alloy microstructures,” International Journal of Plasticity **101**, 1–23 (2018).

[37] D. L. McDowell, “Multiscale crystalline plasticity for materials design,” in *Computational materials system design* (Springer, 2017), pp. 105–146.

[38] G. M. Castelluccio and D. L. McDowell, “A mesoscale approach for growth of 3d microstructurally small fatigue cracks in polycrystals,” International Journal of Damage Mechanics **23**, 791–818 (2014).

[39] V. Prithivirajan, P. Ravi, D. Naragani, and M. D. Sangid, “Direct comparison of microstructure-sensitive fatigue crack initiation via crystal plasticity simulations and in situ high-energy x-ray experiments,” Materials & Design **197**, 109216 (2021).

[40] A. Fatemi and D. F. Socie, “A critical plane approach to multiaxial fatigue damage including out-of-phase loading,” Fatigue & Fracture of Engineering materials & structures **11**, 149–165 (1988).

[41] Y. Murakami, E. T., E. M. K.J., and de los Rios E.R., “The behavior of short cracks,” EGF publication 1 (1986).

[42] M. Nakatani, H. Masuo, Y. Tanaka, and Y. Murakami, “Effect of surface roughness on fatigue strength of ti-6al-4v alloy manufactured by additive manufacturing,” Procedia Structural Integrity **19**, 294–301 (2019).

[43] Y. Yamashita, T. Murakami, R. Mihara, M. Okada, and Y. Murakami, “Defect analysis and fatigue design basis for ni-based superalloy 718 manufactured by additive manufacturing,” Procedia structural integrity **7**, 11–18 (2017).

[44] J. Maierhofer, R. Pippan, and H.-P. Gänser, “Modified nasgro equation for physically short cracks,” International Journal of fatigue **59**, 200–207 (2014).

[45] H. Masuo, Y. Tanaka, S. Morokoshi, H. Yagura, T. Uchida, Y. Yamamoto, and Y. Murakami, “Effects of defects, surface roughness and hip on fatigue strength of ti-6al-4v manufactured by additive manufacturing,” Procedia Structural Integrity **7**, 19–26 (2017).

[46] J. Günther, D. Krewerth, T. Lippmann, S. Leuders, T. Tröster, A. Weidner, H. Biermann, and T. Niendorf, “Fatigue life of additively manufactured ti–6al–4v in the very high cycle fatigue regime,” International Journal of Fatigue **94**, 236–245 (2017).

[47] M. El Haddad, K. Smith, and T. Topper, “Fatigue crack propagation of short cracks,” (1979).

[48] M. Komijani, R. Gracie, and E. Sarvaramini, “Simulation of induced acoustic emission in fractured porous media,” Engineering Fracture Mechanics **210**, 113–131 (2019).

[49] X. Niu, S.-P. Zhu, J.-C. He, D. Liao, J. A. Correia, F. Berto, and Q. Wang, “Defect tolerant fatigue assessment of am materials: size effect and probabilistic prospects,” International Journal of Fatigue **160**, 106884 (2022).

[50] D. McDowell, K. Gall, M. Horstemeyer, and J. Fan, “Microstructure-based fatigue modeling of cast a356-t6 alloy,” Engineering Fracture Mechanics **70**, 49–80 (2003).

[51] Y. Xue, M. Horstemeyer, D. McDowell, H. El Kadiri, and J. Fan, “Microstructure-based multistage fatigue modeling of a cast ae44 magnesium alloy,” International journal of fatigue **29**, 666–676 (2007).

[52] C. Sandoval, L. Malcher, F. Canut, L. Araújo, T. Doca, and J. Araújo, “Micromechanical gurson-based continuum damage under the context of fretting fatigue: influence of the plastic strain field,” International Journal of Plasticity **125**, 235–264 (2020).

[53] A. Tridello, C. Boursier Niutta, F. Berto, G. Qian, and D. S. Paolino, “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**, 1944–1960 (2021).

[54] R. S. Haridas, S. Thapliyal, P. Agrawal, and R. S. Mishra, “Defect-based probabilistic fatigue life estimation model for an additively manufactured aluminum alloy,” Materials Science and Engineering: A **798**, 140082 (2020).

[55] H. Javidrad, B. Koc, H. Bayraktar, U. Simsek, and K. Gunaydin, “Fatigue performance of metal additive manufacturing: a comprehensive overview,” Virtual and Physical Prototyping **19**, e2302556 (2024).

[56] Y. Jia, R. Fu, C. Ling, Z. Shen, L. Zheng, Z. Zhong, and Y. Hong, “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 (2023).

[57] H. Bao, S. Wu, Z. Wu, G. Kang, X. Peng, and P. J. Withers, “A machine-learning fatigue life prediction approach of additively manufactured metals,” Engineering Fracture Mechanics **242**, 107508 (2021).

[58] E. Salvati, A. Tognan, L. Laurenti, M. Pelegatti, and F. De Bona, “A defect-based physics-informed machine learning framework for fatigue finite life prediction in additive manufacturing,” Materials & Design **222**, 111089 (2022).

[59] T. Shi, J. Sun, J. Li, G. Qian, and Y. Hong, “Machine learning based very-high-cycle fatigue life prediction of alsi10mg alloy fabricated by selective laser melting,” International Journal of Fatigue **171**, 107585 (2023).

[60] J. Horňas, J. Běhal, P. Homola, R. Doubrava, M. Holzleitner, and S. Senck, “A machine learning based approach with an augmented dataset for fatigue life prediction of additively manufactured ti-6al-4v samples,” Engineering Fracture Mechanics, 109709 (2023).

[61] A. Senanayaka, W. Tian, T. Falls, and L. Bian, “Understanding the effects of process conditions on thermal–defect relationship: a transfer machine learning approach,” Journal of Manufacturing Science and Engineering **145**, 071010 (2023).

[62] S. Thrun and L. Pratt, “Learning to learn: introduction and overview,” in *Learning to learn* (Springer, 1998), pp. 3–17.

[63] Y. Li, P. Wei, G. Xiang, C. Jia, and H. Liu, “Gear contact fatigue life prediction based on transfer learning,” International Journal of Fatigue **173**, 107686 (2023).

[64] X. Wei, C. Zhang, S. Han, Z. Jia, C. Wang, and W. Xu, “High cycle fatigue sn curve prediction of steels based on transfer learning guided long short term memory network,” International Journal of Fatigue **163**, 107050 (2022).

[65] S. Dong, J. Xiao, X. Hu, N. Fang, L. Liu, and J. Yao, “Deep transfer learning based on bi-lstm and attention for remaining useful life prediction of rolling bearing,” Reliability Engineering & System Safety **230**, 108914 (2023).

[66] L. Xiao, X. Xue, N. Wang, Q. Ren, J. Hua, and F. Wang, “A transfer learning-based approach to fatigue life prediction of corroded bimetallic steel bars using small samples,” Construction and Building Materials **400**, 132679 (2023).

[67] P. Foti, N. Razavi, A. Fatemi, and F. Berto, “Multiaxial fatigue of additively manufactured metallic components: a review of the failure mechanisms and fatigue life prediction methodologies,” Progress in Materials Science **137**, 101126 (2023).

[68] K. Smith, “A stress-strain function for the fatigue of metals,” Journal of materials **5**, 767–778 (1970).

[69] N. Sanaei and A. Fatemi, “Defect-based fatigue life prediction of l-pbf additive manufactured metals,” Engineering Fracture Mechanics **244**, 107541 (2021).

[70] P. Lopez-Crespo, B. Moreno, A. Lopez-Moreno, and J. Zapatero, “Study of crack orientation and fatigue life prediction in biaxial fatigue with critical plane models,” Engineering Fracture Mechanics **136**, 115–130 (2015).

[71] A. Karolczuk and E. Macha, “A review of critical plane orientations in multiaxial fatigue failure criteria of metallic materials,” International Journal of Fracture **134**, 267–304 (2005).

[72] J. Yang, G. Kang, Y. Liu, and Q. Kan, “A novel method of multiaxial fatigue life prediction based on deep learning,” International Journal of Fatigue **151**, 106356 (2021).

[73] K. Pałczyński, D. Skibicki, Ł. Pejkowski, and T. Andrysiak, “Application of machine learning methods in multiaxial fatigue life prediction,” Fatigue & Fracture of Engineering Materials & Structures **46**, 416–432 (2023).

[74] J. Gao, F. Heng, Y. Yuan, and Y. Liu, “A novel machine learning method for multiaxial fatigue life prediction: improved adaptive neuro-fuzzy inference system,” International Journal of Fatigue **178**, 108007 (2024).

[75] S. Chen, Y. Bai, X. Zhou, and A. Yang, “A deep learning dataset for metal multiaxial fatigue life prediction,” Scientific Data **11**, 1027 (2024).

[76] Z. Zheng, X. Li, T. Sun, Z. Huang, and C. Xie, “Multiaxial fatigue life prediction of metals considering loading paths by image recognition and machine learning,” Engineering Failure Analysis **143**, 106851 (2023).

[77] J. Yang, G. Kang, and Q. Kan, “A novel deep learning approach of multiaxial fatigue life-prediction with a self-attention mechanism characterizing the effects of loading history and varying temperature,” International Journal of Fatigue **162**, 106851 (2022).

[78] G. He, Y. Zhao, and C. Yan, “Application of tabular data synthesis using generative adversarial networks on machine learning-based multiaxial fatigue life prediction,” International Journal of Pressure Vessels and Piping **199**, 104779 (2022).

[79] J. Tang, J. Zhou, and Z. C. Tan, “A novel multiaxial fatigue life prediction model based on the critical plane theory and machine-learning method,” The Journal of Strain Analysis for Engineering Design **59**, 139–149 (2024).

[80] A. Karolczuk, D. Skibicki, and Ł. Pejkowski, “Gaussian process for machine learning-based fatigue life prediction model under multiaxial stress–strain conditions,” Materials **15**, 7797 (2022).

[81] X. Sun, K. Zhou, S. Shi, K. Song, and X. Chen, “A new cyclical generative adversarial network based data augmentation method for multiaxial fatigue life prediction,” International Journal of Fatigue **162**, 106996 (2022).

[82] P. Zhang, K. Tang, A. Wang, H. Wu, and Z. Zhong, “Neural network integrated with symbolic regression for multiaxial fatigue life prediction,” International Journal of Fatigue **188**, 108535 (2024).

[83] H. Liao, X. Sun, and X. Chen, “A physics-informed residual neural network in the meta-learning framework with non-proportionality measurement for multiaxial fatigue life prediction,” International Journal of Fatigue, 109336 (2025).

[84] M. Bartošák, J. Halamka, L. Beránek, M. Koukolikova, M. Slanỳ, M. Pagáč, and J. Džugan, “Using physics-informed neural networks to predict the lifetime of laser powder bed fusion

processed 316l stainless steel under multiaxial low-cycle fatigue loading," International Journal of Fatigue **190**, 108608 (2025).

[85] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings, et al., "Advances and open problems in federated learning," Foundations and trends® in machine learning **14**, 1–210 (2021).

[86] E. T. M. Beltrán, M. Q. Pérez, P. M. S. Sánchez, S. L. Bernal, G. Bovet, M. G. Pérez, G. M. Pérez, and A. H. Celdrán, "Decentralized federated learning: fundamentals, state of the art, frameworks, trends, and challenges," IEEE Communications Surveys & Tutorials **25**, 2983–3013 (2023).

[87] L. Yuan, Z. Wang, L. Sun, P. S. Yu, and C. G. Brinton, "Decentralized federated learning: a survey and perspective," IEEE Internet of Things Journal **11**, 34617–34638 (2024).

[88] S. Kalra, J. Wen, J. C. Cresswell, M. Volkovs, and H. R. Tizhoosh, "Decentralized federated learning through proxy model sharing," Nature communications **14**, 2899 (2023).

[89] M. Ye, X. Fang, B. Du, P. C. Yuen, and D. Tao, "Heterogeneous federated learning: state-of-the-art and research challenges," ACM Computing Surveys **56**, 1–44 (2023).

[90] A. Rakotomamonjy, M. Vono, H. J. M. Ruiz, and L. Ralaivola, "Personalised federated learning on heterogeneous feature spaces," Transactions on Machine Learning Research (2024).

[91] Y. Liu, Y. Kang, T. Zou, Y. Pu, Y. He, X. Ye, Y. Ouyang, Y.-Q. Zhang, and Q. Yang, "Vertical federated learning: concepts, advances, and challenges," IEEE transactions on knowledge and data engineering **36**, 3615–3634 (2024).

[92] A. Arunan, Y. Qin, X. Li, and C. Yuen, "A federated learning-based industrial health prognostics for heterogeneous edge devices using matched feature extraction," IEEE Transactions on Automation Science and Engineering **21**, 3065–3079 (2023).

[93] D. van Esch, E. Sarbar, T. Lucassen, J. O'Brien, T. Breiner, M. Prasad, E. Crew, C. Nguyen, and F. Beaufays, "Writing across the world's languages: deep internationalization for gboard, the google keyboard," arXiv preprint arXiv:1912.01218 (2019).

[94] A. Hard, K. Rao, R. Mathews, S. Ramaswamy, F. Beaufays, S. Augenstein, H. Eichner, C. Kiddon, and D. Ramage, "Federated learning for mobile keyboard prediction," arXiv preprint arXiv:1811.03604 (2018).

[95] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "FedProx," MLSys (2020).

[96] Z. Ling, Z. Yue, J. Xia, T. Wang, M. Chen, and X. Lian, "Fedentropy: efficient federated learning for non-iid scenarios using maximum entropy judgment-based client selection," in 2023 ieee intl conf on parallel & distributed processing with applications, big data & cloud

computing, sustainable computing & communications, social computing & networking (ispa/bdcloud/socialcom/sustaincom) (IEEE, 2023), pp. 56–63.

[97] T. Li, S. Hu, A. Beirami, and V. Smith, "Ditto: fair and robust federated learning through personalization," in International conference on machine learning (PMLR, 2021), pp. 6357–6368.

[98] V. Smith, C.-K. Chiang, M. Sanjabi, and A. S. Talwalkar, "Federated multi-task learning," Advances in neural information processing systems **30** (2017).

[99] D. Li and J. Wang, "Fedmd: heterogenous federated learning via model distillation," arXiv preprint arXiv:1910.03581 (2019).

[100] M. Mehta and C. Shao, "A greedy agglomerative framework for clustered federated learning," IEEE Transactions on Industrial Informatics **19**, 11856–11867 (2023).

[101] D. Gao, X. Yao, and Q. Yang, "A survey on heterogeneous federated learning," arXiv preprint arXiv:2210.04505 (2022).

[102] A. Khan, M. ten Thij, and A. Wilbik, "Vertical federated learning: a structured literature review," Knowledge and Information Systems, 1–39 (2025).

[103] I. Ceballos, V. Sharma, E. Mugica, A. Singh, A. Roman, P. Vepakomma, and R. Raskar, "Splitnn-driven vertical partitioning," arXiv preprint arXiv:2008.04137 (2020).

[104] T. Zou, Y. Liu, Y. Kang, W. Liu, Y. He, Z. Yi, Q. Yang, and Y.-Q. Zhang, "Defending batch-level label inference and replacement attacks in vertical federated learning," IEEE Transactions on Big Data **10**, 1016–1027 (2022).

[105] A. Li, A. Poudel, S. Shao, N. Shamsaei, and J. Liu, "Nondestructive fatigue life prediction for additively manufactured metal parts through a multimodal transfer learning framework," IISE Transactions **57**, 1344–1359 (2025).

[106] S. Saha and T. Ahmad, "Federated transfer learning: concept and applications," Intelligenza Artificiale **15**, 35–44 (2021).

[107] Y. Liu, Y. Kang, C. Xing, T. Chen, and Q. Yang, "A secure federated transfer learning framework," IEEE Intelligent Systems **35**, 70–82 (2020).

[108] S. Feng, B. Li, H. Yu, Y. Liu, and Q. Yang, "Semi-supervised federated heterogeneous transfer learning," Knowledge-Based Systems **252**, 109384 (2022).

[109] Y. Kang, Y. He, J. Luo, T. Fan, Y. Liu, and Q. Yang, "Privacy-preserving federated adversarial domain adaptation over feature groups for interpretability," IEEE Transactions on Big Data **10**, 879–890 (2022).

[110] S. Feng and H. Yu, "Multi-participant multi-class vertical federated learning," arXiv preprint arXiv:2001.11154 (2020).

[111] N. Shi and R. A. Kontar, "Personalized federated learning via domain adaptation with an application to distributed 3d printing," Technometrics **65**, 328–339 (2023).

[112] X. Yang, B. Niu, T. Lan, and C. Zhang, "Federated multi-task bayesian network learning in the presence of overlapping and distinct variables," IISE Transactions **57**, 773–787 (2025).

[113] K. Oishi, Y. Sei, Y. Tahara, and A. Ohsuga, "Federated learning algorithm handling missing attributes," in 2023 ieee international conference on internet of things and intelligence systems (iotais) (IEEE, 2023), pp. 146–151.

[114] D. Adhikari, W. Jiang, J. Zhan, Z. He, D. B. Rawat, U. Aickelin, and H. A. Khorshidi, "A comprehensive survey on imputation of missing data in internet of things," ACM Computing Surveys **55**, 1–38 (2022).

[115] S. Min, H. Asif, X. Wang, and J. Vaidya, "Cafe: improved federated data imputation by leveraging missing data heterogeneity," IEEE Transactions on Knowledge and Data Engineering (2025).

[116] P. Li, H. Zhang, Y. Wu, L. Qian, R. Yu, D. Niyato, and X. Shen, "Filling the missing: exploring generative ai for enhanced federated learning over heterogeneous mobile edge devices," IEEE Transactions on Mobile Computing **23**, 10001–10015 (2024).

[117] P. J. Maliakel, S. Ilager, and I. Brandic, "Fligan: enhancing federated learning with incomplete data using gan," in Proceedings of the 7th international workshop on edge systems, analytics and networking (2024), pp. 1–6.

[118] B. Sun and K. Saenko, "Deep coral: correlation alignment for deep domain adaptation," in European conference on computer vision (Springer, 2016), pp. 443–450.

[119] X. Chen, S. Wang, J. Wang, and M. Long, "Representation subspace distance for domain adaptation regression.," in Icml (2021), pp. 1749–1759.

[120] M. Dhaini, M. Berar, P. Honeine, and A. Van Exem, "Unsupervised domain adaptation for regression using dictionary learning," Knowledge-Based Systems **267**, 110439 (2023).

[121] A. Farahani, S. Voghoei, K. Rasheed, and H. R. Arabnia, "A brief review of domain adaptation," Advances in data science and information engineering: proceedings from ICDATA 2020 and IKE 2020, 877–894 (2021).

[122] N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy, "Optimal transport for domain adaptation," IEEE transactions on pattern analysis and machine intelligence **39**, 1853–1865 (2016).

[123] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-efficient learning of deep networks from decentralized data,” in Artificial intelligence and statistics (PMLR, 2017), pp. 1273–1282.

[124] F. Farnia, A. Reisizadeh, R. Pedarsani, and A. Jadbabaie, “An optimal transport approach to personalized federated learning,” IEEE Journal on Selected Areas in Information Theory **3**, 162–171 (2022).

[125] H. Li, W. Huang, J. Wang, and Y. Shi, “Global and local prompts cooperation via optimal transport for federated learning,” in Proceedings of the ieee/cvf conference on computer vision and pattern recognition (2024), pp. 12151–12161.

[126] S. Ek, K. Wang, F. Portet, P. Lalanda, and J. Cao, “Fedali: personalized federated learning with aligned prototypes through optimal transport,” arXiv preprint arXiv:2411.10595 (2024).

[127] L. M. Pereira and M. H. Amini, “A survey on optimal transport for machine learning: theory and applications,” IEEE Access (2025).

[128] G. Peyré, M. Cuturi, et al., “Computational optimal transport: with applications to data science,” Foundations and Trends® in Machine Learning **11**, 355–607 (2019).

[129] R. Chhaibi, S. Gratton, and S. Vaiter, “Faster computation of entropic optimal transport via stable low frequency modes,” arXiv preprint arXiv:2506.14780 (2025).

[130] M. Cuturi, “Sinkhorn distances: lightspeed computation of optimal transport,” Advances in neural information processing systems **26** (2013).

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

[132] B. B. Damodaran, B. Kellenberger, R. Flamary, D. Tuia, and N. Courty, “Deepjdot: deep joint distribution optimal transport for unsupervised domain adaptation,” in Proceedings of the european conference on computer vision (eccv) (2018), pp. 447–463.

[133] S. K. Everton, M. Hirsch, P. Stravroulakis, R. K. Leach, and A. T. Clare, “Review of in-situ process monitoring and in-situ metrology for metal additive manufacturing,” Materials & Design **95**, 431–445 (2016).

[134] P. K. Rao, J. Liu, D. Roberson, Z. Kong, and C. Williams, “Online real-time quality monitoring in additive manufacturing processes using heterogeneous sensors,” Journal of Manufacturing Science and Engineering **137**, 061007 (2015).

[135] S. Ford and M. Despeisse, “Additive manufacturing and sustainability: an exploratory study of the advantages and challenges,” Journal of cleaner Production **137**, 1573–1587 (2016).

[136] N. Guo and M. C. Leu, “Additive manufacturing: technology, applications and research needs,” Frontiers of mechanical engineering **8**, 215–243 (2013).

[137] N. Shamsaei, A. Yadollahi, L. Bian, and S. M. Thompson, “An overview of direct laser deposition for additive manufacturing; part ii: mechanical behavior, process parameter optimization and control,” Additive manufacturing **8**, 12–35 (2015).

[138] R. DeVor, S. Kapoor, J. Cao, and K. Ehmann, “Transforming the landscape of manufacturing: distributed manufacturing based on desktop manufacturing (dm) 2,” Journal of manufacturing science and engineering **134**, 041004 (2012).

[139] A. Ghobadian, I. Talavera, A. Bhattacharya, V. Kumar, J. A. Garza-Reyes, and N. O’regan, “Examining legitimatisation of additive manufacturing in the interplay between innovation, lean manufacturing and sustainability,” International Journal of Production Economics **219**, 457–468 (2020).

[140] J. Liu, J. Ye, D. Silva Izquierdo, A. Vinel, N. Shamsaei, and S. Shao, “A review of machine learning techniques for process and performance optimization in laser beam powder bed fusion additive manufacturing,” Journal of Intelligent Manufacturing **34**, 3249–3275 (2023).

[141] A. Yadollahi and N. Shamsaei, “Additive manufacturing of fatigue resistant materials: challenges and opportunities,” International Journal of Fatigue **98**, 14–31 (2017).

[142] R. Cunningham, S. P. Narra, C. Montgomery, J. Beuth, and A. Rollett, “Synchrotron-based x-ray microtomography characterization of the effect of processing variables on porosity formation in laser power-bed additive manufacturing of ti-6al-4v,” Jom **69**, 479–484 (2017).

[143] A. J. Sterling, B. Torries, N. Shamsaei, S. M. Thompson, and D. W. Seely, “Fatigue behavior and failure mechanisms of direct laser deposited ti–6al–4v,” Materials Science and Engineering: A **655**, 100–112 (2016).

[144] M. Seifi, M. Gorelik, J. Waller, N. Hrabe, N. Shamsaei, S. Daniewicz, and J. J. Lewandowski, “Progress towards metal additive manufacturing standardization to support qualification and certification,” Jom **69**, 439–455 (2017).

[145] R. Shrestha, N. Shamsaei, M. Seifi, and N. Phan, “An investigation into specimen property to part performance relationships for laser beam powder bed fusion additive manufacturing,” Additive Manufacturing **29**, 100807 (2019).

[146] A. Fatemi, R. Molaei, J. Simsiriwong, N. Sanaei, J. Pegues, B. Torries, N. Phan, and N. Shamsaei, “Fatigue behaviour of additive manufactured materials: an overview of some recent experimental studies on ti-6al-4v considering various processing and loading direction effects,” Fatigue & fracture of engineering materials & structures **42**, 991–1009 (2019).

[147] R. Molaei, A. Fatemi, N. Sanaei, J. Pegues, N. Shamsaei, S. Shao, P. Li, D. Warner, and N. Phan, “Fatigue of additive manufactured ti-6al-4v, part ii: the relationship between microstructure, material cyclic properties, and component performance,” International Journal of Fatigue **132**, 105363 (2020).

[148] J. Pegues, S. Shao, N. Shamsaei, N. Sanaei, A. Fatemi, D. Warner, P. Li, and N. Phan, “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 (2020).

[149] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*, Vol. 4, 4 (Springer, 2006).

[150] Y. Mingqiang, K. Kidiyo, R. Joseph, et al., “A survey of shape feature extraction techniques,” Pattern recognition **15**, 43–90 (2008).

[151] P. Nezhadfar, E. Burford, K. Anderson-Wedge, B. Zhang, S. Shao, S. Daniewicz, and N. Shamsaei, “Fatigue crack growth behavior of additively manufactured 17-4 ph stainless steel: effects of build orientation and microstructure,” International Journal of Fatigue **123**, 168–179 (2019).

[152] P. Nezhadfar, R. Shrestha, N. Phan, and N. Shamsaei, “Fatigue data for laser beam powder bed fused 17-4 ph stainless steel specimens in different heat treatment and surface roughness conditions,” Data in brief **25**, 104215 (2019).

[153] A. Li, A. Oladazimi, M. Muhammad, S. Soman, S. Shao, N. Shamsaei, and J. Liu, “Curvature-based defect features for nondestructive fatigue life prediction in laser powder bed fusion,” Progress in Additive Manufacturing, 1–16 (2026).

[154] L. P. Eisenhart, *A treatise on the differential geometry of curves and surfaces* (Courier Corporation, 2013).

[155] N. Altman and M. Krzywinski, “Points of significance: association, correlation and causation.,” Nature methods **12** (2015).

[156] D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti, “Detecting novel associations in large data sets,” science **334**, 1518–1524 (2011).

[157] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik, “Support vector regression machines,” Advances in neural information processing systems **9** (1996).

[158] M. Awad and R. Khanna, “Support vector regression,” in *Efficient learning machines: theories, concepts, and applications for engineers and system designers* (Springer, 2015), pp. 67–80.

[159] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," Statistics and computing **14**, 199–222 (2004).

[160] M. T. Ribeiro, S. Singh, and C. Guestrin, "Model-agnostic interpretability of machine learning," arXiv preprint arXiv:1606.05386 (2016).

[161] P. Nezhadfar, R. Shrestha, N. Phan, and N. Shamsaei, "Fatigue data for laser beam powder bed fused 17-4 ph stainless steel specimens in different heat treatment and surface roughness conditions," Data in brief **25**, 104215 (2019).

[162] A. Soltani-Tehrani, J. Pegues, and N. Shamsaei, "Fatigue behavior of additively manufactured 17-4 ph stainless steel: the effects of part location and powder re-use," Additive Manufacturing **36**, 101398 (2020).

[163] J. Ye, "Machine learning-based additive manufacturing process optimization and quality prediction with photomicrography and tomography," Available at `https://etd.auburn.edu/handle/10415/8912`, PhD dissertation (Auburn university, Auburn, AL, Aug. 2023).

[164] M. D. Abràmoff, P. J. Magalhães, and S. J. Ram, "Image processing with imagej," Biophotonics international **11**, 36–42 (2004).

[165] M. Musy, G. Jacquenot, G. Dalmasso, R. de Bruin, A. Pollack, F. Claudi, C. Badger, B. Sullivan, D. Hrisca, D. Volpatto, et al., "Vedo, a python module for scientific analysis and visualization of 3d objects and point clouds," Zenodo **1** (2021).

[166] B. OH, "The exponential law of endurance tests," in Proc am soc test mater, Vol. 10 (1910), pp. 625–630.

[167] S. Kwofie and H. Chandler, "Fatigue life prediction under conditions where cyclic creep–fatigue interaction occurs," International Journal of Fatigue **29**, 2117–2124 (2007).

[168] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," Advances in neural information processing systems **30** (2017).

[169] J. Liu, C. Liu, Y. Bai, P. Rao, C. B. Williams, and Z. Kong, "Layer-wise spatial modeling of porosity in additive manufacturing," IISE Transactions **51**, 109–123 (2019).

[170] X. Peng, S. Wu, W. Qian, J. Bao, Y. Hu, Z. Zhan, G. Guo, and P. J. Withers, "The potency of defects on fatigue of additively manufactured metals," International Journal of Mechanical Sciences **221**, 107185 (2022).

[171] F. H. Kim, A. L. Pintar, S. P. Moylan, and E. J. Garboczi, "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**, 111002 (2019).

[172] G. Meneghetti, D. Rigon, and C. Gennari, “An analysis of defects influence on axial fatigue strength of maraging steel specimens produced by additive manufacturing,” International journal of fatigue **118**, 54–64 (2019).

[173] R. Xu, S. Huang, Z. Song, Y. Gao, and J. Wu, “A deep mixed-effects modeling approach for real-time monitoring of metal additive manufacturing process,” IISE Transactions, 1–15 (2023).

[174] Y. Hu, S. Wu, P. Withers, J. Zhang, H. Bao, Y. Fu, and G. Kang, “The effect of manufacturing defects on the fatigue life of selective laser melted ti-6al-4v structures,” Materials & Design **192**, 108708 (2020).

[175] S. Shao, M. J. Mahtabi, N. Shamsaei, and S. M. Thompson, “Solubility of argon in laser additive manufactured α-titanium under hot isostatic pressing condition,” Computational Materials Science **131**, 209–219 (2017).

[176] Z. Qin, N. Kang, F. Zhang, Z. Wang, Q. Wang, J. Chen, X. Lin, and W. Huang, “Role of defects on the high cycle fatigue behavior of selective laser melted al–mg–sc–zr alloy,” International Journal of Fracture, 1–15 (2021).

[177] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, “Nih image to imagej: 25 years of image analysis,” Nature methods **9**, 671–675 (2012).

[178] A. A693-16, *Standard specification for precipitation-hardening stainless and heat-resisting steel plate, sheet, and strip*, 2016.

[179] C.-W. Hang and M. P. Singh, “Trust-based recommendation based on graph similarity,” in Proceedings of the 13th international workshop on trust in agent societies (trust). toronto, canada, Vol. 82 (2010).

[180] L. Yao, C. Mao, and Y. Luo, “Graph convolutional networks for text classification,” in Proceedings of the aaai conference on artificial intelligence, Vol. 33, 01 (2019), pp. 7370–7377.

[181] T.-A. Song, S. R. Chowdhury, F. Yang, H. Jacobs, G. El Fakhri, Q. Li, K. Johnson, and J. Dutta, “Graph convolutional neural networks for alzheimer’s disease classification,” in 2019 ieee 16th international symposium on biomedical imaging (isbi 2019) (IEEE, 2019), pp. 414–417.

[182] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” arXiv preprint arXiv:1609.02907 (2016).

[183] D. P. Kingma and J. Ba, “Adam: a method for stochastic optimization,” arXiv preprint arXiv:1412.6980 (2014).

[184] L. Wang, S.-P. Zhu, C. Luo, X. Niu, and J.-C. He, “Defect driven physics-informed neural network framework for fatigue life prediction of additively manufactured materials,” Philosophical Transactions of the Royal Society A **381**, 20220386 (2023).

[185] S. Inayathullah and R. Buddala, “Review of machine learning applications in additive manufacturing,” Results in Engineering **25**, 103676 (2025).

[186] F. Haertel, J. Mangler, N. Klievtsova, C. Mader, E. Rigger, and S. Rinderle-Ma, “Privacy and confidentiality requirements engineering for process data,” in International conference on business process management (Springer, 2025), pp. 216–231.

[187] I. Jussen, F. Möller, J. Schweihoff, A. Gieß, G. Giussani, and B. Otto, “Issues in inter-organizational data sharing: findings from practice and research challenges,” Data & Knowledge Engineering **150**, 102280 (2024).

[188] M. Mehta, M. V. Bimrose, D. J. Mcgregor, W. P. King, and C. Shao, “Federated learning enables privacy-preserving and data-efficient dimension prediction and part qualification across additive manufacturing factories,” Journal of Manufacturing Systems **74**, 752–761 (2024).

[189] S. Chowdhury, N. Yadaiah, C. Prakash, S. Ramakrishna, S. Dixit, L. R. Gupta, and D. Buddhi, “Laser powder bed fusion: a state-of-the-art review of the technology, materials, properties & defects, and numerical modelling,” Journal of Materials Research and Technology **20**, 2109–2172 (2022).

[190] T. Gao, A. Li, X. Zhang, G. Harris, and J. Liu, “A data-driven process-quality-property analytical framework for conductive composites in additive manufacturing,” Manufacturing Letters **35**, 626–635 (2023).

[191] H.-C. Tran, Y.-L. Lo, H.-C. Yang, H.-C. Hsiao, F.-T. Cheng, and T.-H. Kuo, “Intelligent additive manufacturing architecture for enhancing uniformity of surface roughness and mechanical properties of laser powder bed fusion components,” IEEE Transactions on Automation Science and Engineering (2022).

[192] R. Ibrahim, “Overview of structural life assessment and reliability, part ii: fatigue life and reliability assessment of naval ship structures,” Journal of Ship Production and Design **31**, 100–128 (2015).

[193] S. Han, T. D. Dinh, I. De Baere, M. Boone, I. Josipovic, and W. Van Paepegem, “Study of the effect of defects on fatigue life prediction of additive manufactured ti-6al-4v by combined use of micro-computed tomography and fracture-mechanics-based simulation,” International Journal of Fatigue **180**, 108110 (2024).

[194] X. Chen, B. Zheng, and H. Liu, “Optical and digital microscopic imaging techniques and applications in pathology,” Analytical Cellular Pathology **34**, 5–18 (2011).

[195] N. Shi, F. Lai, R. Al Kontar, and M. Chowdhury, “Fed-ensemble: ensemble models in federated learning for improved generalization and uncertainty quantification,” IEEE Transactions on Automation Science and Engineering **21**, 2792–2803 (2023).

[196] R. Kontar, N. Shi, X. Yue, S. Chung, E. Byon, M. Chowdhury, J. Jin, W. Kontar, N. Masoud, M. Nouiehed, et al., “The internet of federated things (ioft),” IEEE Access **9**, 156071–156113 (2021).

[197] Q. Fang, G. Xiong, M. Zhou, T. S. Tamir, C.-B. Yan, H. Wu, Z. Shen, and F.-Y. Wang, “Process monitoring, diagnosis and control of additive manufacturing,” IEEE Transactions on Automation Science and Engineering **21**, 1041–1067 (2022).

[198] W. Frazier, Y. Lu, and P. Witherell, “Empowering small and medium-sized enterprises through effective additive manufacturing data management,” (2024).

[199] M. Sæterbø and W. D. Solvang, “Metal additive manufacturing adoption in smes: technical attributes, challenges, and opportunities,” Journal of Manufacturing Processes **128**, 175–189 (2024).

[200] P. Kulkarni, A. Kumar, G. Chate, and P. Dandannavar, “Elements of additive manufacturing technology adoption in small-and medium-sized companies,” Innovation & Management Review **18**, 400–416 (2021).

[201] A. Jamwal, R. Agrawal, and M. Sharma, “Challenges and opportunities for manufacturing smes in adopting industry 4.0 technologies for achieving sustainability: empirical evidence from an emerging economy,” Operations Management Research **18**, 718–743 (2025).

[202] G. Kang, L. Jiang, Y. Yang, and A. G. Hauptmann, “Contrastive adaptation network for unsupervised domain adaptation,” in Proceedings of the ieee/cvf conference on computer vision and pattern recognition (2019), pp. 4893–4902.

[203] A. Genevay, “Entropy-regularized optimal transport for machine learning,” PhD thesis (Université Paris sciences et lettres, 2019).

[204] A. Das, T. Castiglia, S. Wang, and S. Patterson, “Cross-silo federated learning for multi-tier networks with vertical and horizontal data partitioning,” ACM Transactions on Intelligent Systems and Technology (TIST) **13**, 1–27 (2022).

[205] A. El Ouadrhiri and A. Abdelhadi, “Differential privacy for deep and federated learning: a survey,” IEEE access **10**, 22359–22380 (2022).

[206] H. Fang and Q. Qian, “Privacy preserving machine learning with homomorphic encryption and federated learning,” Future Internet **13**, 94 (2021).

[207] Q. Li, Z. Wen, Z. Wu, S. Hu, N. Wang, Y. Li, X. Liu, and B. He, “A survey on federated learning systems: vision, hype and reality for data privacy and protection,” IEEE Transactions on Knowledge and Data Engineering **35**, 3347–3366 (2021).

[208] R. Wu, A. Scaglione, H.-T. Wai, N. Karakoc, K. Hreinsson, and W.-K. Ma, “Federated block coordinate descent scheme for learning global and personalized models,” in Proceedings of the aaai conference on artificial intelligence, Vol. 35, 12 (2021), pp. 10355–10362.

[209] A. Li, J. Liu, S. Shao, and N. Shamsaei, “Defects classification via hierarchical graph convolutional network in l-pbf additive manufacturing,” in 2022 international solid freeform fabrication symposium (2022).

[210] C. R. Brooks and B. McGill, “The application of scanning electron microscopy to fractography,” Materials characterization **33**, 195–243 (1994).

[211] A. Standard, “E8/e8m. standard test methods for tension testing of metallic materials. west conshohocken (pa): astm,” in (International, 2011).

[212] A. Sadeghian and S. Adibnazari, “Design-informed evaluation and fretting damage mapping of additive-manufactured ti-6al-4v components,” Designs **9**, 100 (2025).

[213] M. Muhammad, M. S. Yasin, S. Soman, S. Shao, and N. Shamsaei, “Defect features critical to the fatigue of additively manufactured ti-6al-4v,” Theoretical and Applied Fracture Mechanics, 104981 (2025).

[214] L. Gan, H. Wu, and Z. Zhong, “Multiaxial fatigue life prediction based on a simplified energy-based model,” International Journal of Fatigue **144**, 106036 (2021).

[215] O. Maron and T. Lozano-Pérez, “A framework for multiple-instance learning,” Advances in neural information processing systems **10** (1997).

[216] M.-A. Carbonneau, V. Cheplygina, E. Granger, and G. Gagnon, “Multiple instance learning: a survey of problem characteristics and applications,” Pattern recognition **77**, 329–353 (2018).

[217] M. Ilse, J. Tomczak, and M. Welling, “Attention-based deep multiple instance learning,” in International conference on machine learning (PMLR, 2018), pp. 2127–2136.

[218] W. Qian, S. Wu, Z. Wu, S. Ahmed, W. Zhang, G. Qian, and P. J. Withers, “In situ x-ray imaging of fatigue crack growth from multiple defects in additively manufactured alsi10mg alloy,” International Journal of Fatigue **155**, 106616 (2022).

Appendix 1

Appendix of Chapter 4

A1.1 The Derivation of HGCN Weight Updates

This work presents the derivative of back-propagation to update HGCN model parameters (θ_1 of GCN_1 and θ_2 of GCN_2) as mentioned in Section 4.3.1 via adaptive moment estimation (Adam). For convenience, this work assumes these two GCNs share common hyperparameters (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 $\epsilon = 10^{-8}$ [183]. The parameters are updated based on,

$$\begin{aligned} \theta_1^t &= \theta_1^{t-1} - \alpha \cdot \frac{\hat{\mathbf{m}}_1^t}{\sqrt{\hat{\mathbf{v}}_1^t} + \epsilon} \\ \theta_2^t &= \theta_2^{t-1} - \alpha \cdot \frac{\hat{\mathbf{m}}_2^t}{\sqrt{\hat{\mathbf{v}}_2^t} + \epsilon} \end{aligned} \tag{A1.1}$$

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

$$\begin{aligned} \hat{\mathbf{m}}_1^t &= \frac{\mathbf{m}_1^t}{1-\gamma_1},\ \hat{\mathbf{m}}_2^t = \frac{\mathbf{m}_2^t}{1-\gamma_1} \\ \hat{\mathbf{v}}_1^t &= \frac{\mathbf{v}_1^t}{1-\gamma_2},\ \hat{\mathbf{v}}_2^t = \frac{\mathbf{v}_2^t}{1-\gamma_2} \end{aligned} \tag{A1.2}$$

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

$$\begin{aligned} \mathbf{m}_1^t &= \gamma_1 \cdot \mathbf{m}_1^{t-1} + (1-\gamma_1) \cdot \nabla_{\theta_1} \mathcal{L}^{S,t}(\theta_1^{t-1}, \theta_2^{t-1}) \\ \mathbf{m}_2^t &= \gamma_1 \cdot \mathbf{m}_2^{t-1} + (1-\gamma_1) \cdot \nabla_{\theta_2} \mathcal{L}^{S,t}(\theta_1^{t-1}, \theta_2^{t-1}) \\ \mathbf{v}_1^t &= \gamma_2 \cdot \mathbf{v}_1^{t-1} + (1-\gamma_2) \cdot (\nabla_{\theta_1} \mathcal{L}^{S,t}(\theta_1^{t-1}, \theta_2^{t-1}) \odot \nabla_{\theta_1} \mathcal{L}^{S,t}(\theta_1^{t-1}, \theta_2^{t-1})) \\ \mathbf{v}_2^t &= \gamma_2 \cdot \mathbf{v}_2^{t-1} + (1-\gamma_2) \cdot (\nabla_{\theta_2} \mathcal{L}^{S,t}(\theta_1^{t-1}, \theta_2^{t-1}) \odot \nabla_{\theta_2} \mathcal{L}^{S,t}(\theta_1^{t-1}, \theta_2^{t-1})) \end{aligned} \tag{A1.3}$$

To simplify the derivative of the update for θ_1 and θ_2, this work considers single-layer GCN_1 and GCN_2 since it is simple 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 GCN_1 and GCN_2 as mentioned in Section 4.3.1. This work can rewrite Eq.(5.5) and Eq.(5.6) as $g_1(\theta_1; \mathbf{X}_1^S) = \text{ReLU}(\tilde{\mathbf{A}}_1^* \mathbf{X}_1^S \theta_1)$ and $g^{(2)}(g_2(\theta_2; \mathbf{X}_2^S) = \ddot{s} \circ \mathbf{H}_2 = \ddot{s} \circ \text{ReLU}(\tilde{\mathbf{A}}_2^* [\mathbf{X}_2^S | g_1(\theta_1; \mathbf{X}_1^S)] \theta_2)$. Here, this work uses $\ddot{s}$ to replace Softmax. Now, this work solves the $\nabla_{\theta_1}\mathcal{L}^S(\theta_1,\theta_2)$ and $\nabla_{\theta_2}\mathcal{L}^S(\theta_1,\theta_2)$ ($\mathcal{L}^S$ is formulated in Eq.(5.2) and Eq.(5.7) and this work ignores the subscript $t-1$),

$$
\begin{aligned}
\nabla_{\theta_1}\mathcal{L}^S(\theta_1,\theta_2) &= \frac{\partial \mathcal{L}^S}{\partial \theta_1} \\
&= \frac{\partial \mathcal{L}^S}{\partial g_2(\theta_2)} \cdot \frac{\partial g_2(\theta_2)}{\partial g_1(\theta_1)} \cdot \frac{\partial g_1(\theta_1)}{\partial \theta_1} \\
&= -\frac{O^S}{g_2(\theta_2)} \cdot \nabla_{\mathbf{H}_2}\ddot{s} \cdot (\theta_2^{\mathsf{T}}\tilde{\mathbf{A}}_2^*) \cdot [\mathbf{0}|\mathbf{I}] \cdot (\tilde{\mathbf{A}}_1^*\mathbf{X}_1^S) \\
&= -\frac{O^S}{\ddot{s}(\mathbf{H}_2)} \cdot \nabla_{\mathbf{H}_2}\ddot{s} \cdot (\theta_2^{\mathsf{T}}\tilde{\mathbf{A}}_2^*) \cdot [\mathbf{0}|\mathbf{I}] \cdot (\tilde{\mathbf{A}}_1^*\mathbf{X}_1^S) \\
&= -O^S \cdot (\mathbf{I} - \ddot{s}(\mathbf{H}_2)) \cdot (\theta_2^{\mathsf{T}}\tilde{\mathbf{A}}_2^*) \cdot [\mathbf{0}|\mathbf{I}] \cdot (\tilde{\mathbf{A}}_1^*\mathbf{X}_1^S) \\
&= -O^S \cdot (\mathbf{I} - \ddot{s}(\mathrm{ReLU}(\tilde{\mathbf{A}}_2^*[\mathbf{X}_2^S|g_1(\theta_1;\mathbf{X}_1^S)]\theta_2)) \\
&\cdot (\theta_2^{\mathsf{T}}\tilde{\mathbf{A}}_2^*) \cdot [\mathbf{0}|\mathbf{I}] \cdot (\tilde{\mathbf{A}}_1^*\mathbf{X}_1^S) \\
&= -(\underbrace{\mathbf{I} - \ddot{s}(\mathrm{ReLU}(\tilde{\mathbf{A}}_2^*[\mathbf{X}_2^S|\mathrm{ReLU}(\tilde{\mathbf{A}}_1^*\mathbf{X}_1^S}_{\text{Internal constant part}} \underbrace{\theta_1)]\theta_2)))}_{\text{Parameter part}} \\
&\cdot (\theta_2^{\mathsf{T}} \otimes \underbrace{\tilde{\mathbf{A}}_2^*)O^S \cdot [\mathbf{0}|\mathbf{I}] \cdot (\tilde{\mathbf{A}}_1^*\mathbf{X}_1^S)}_{\text{External constant part}}
\end{aligned}
\tag{A1.4}
$$

$$
\begin{aligned}
\nabla_{\theta_2}\mathcal{L}^S(\theta_1,\theta_2) &= \frac{\partial \mathcal{L}^S}{\partial \theta_2} \\
&= \frac{\partial \mathcal{L}^S}{\partial g_2(\theta_2)} \cdot \frac{\partial g_2(\theta_2)}{\partial \theta_2} \\
&= -\frac{O^S}{g_2}(\theta_2) \cdot \nabla_{\mathbf{H}_2}\ddot{s} \cdot (\tilde{\mathbf{A}}_2^*[\mathbf{X}_2^S|g_1(\theta_1;\mathbf{X}_1^S)]) \\
&= -\frac{O^S}{\ddot{s}(\mathbf{H}_1)} \cdot \nabla_{\mathbf{H}_2}\ddot{s} \cdot (\tilde{\mathbf{A}}_2^*[\mathbf{X}_2^S|g_1(\theta_1;\mathbf{X}_1^S)]) \\
&= -O^S(\mathbf{I} - \ddot{s}(\mathbf{H}_2)) \cdot (\tilde{\mathbf{A}}_2^*[\mathbf{X}_2^S|g_1(\theta_1;\mathbf{X}_1^S)]) \\
&= -O^S(\mathbf{I} - \ddot{s}(\mathrm{ReLU}(\tilde{\mathbf{A}}_2^*[\mathbf{X}_2^S|g_1(\theta_1;\mathbf{X}_1^S)]\theta_2))) \\
&\cdot (\tilde{\mathbf{A}}_2^*[\mathbf{X}_2^S|g_1(\theta_1;\mathbf{X}_1^S)]) \\
&= -O^S(\mathbf{I} - \ddot{s}(\mathrm{ReLU}(\tilde{\mathbf{A}}_2^*[\mathbf{X}_2^S|\mathrm{ReLU}(\tilde{\mathbf{A}}_1^*\mathbf{X}_1^S\theta_1)]\theta_2))) \\
&\cdot (\tilde{\mathbf{A}}_2^*[\mathbf{X}_2^S|\mathrm{ReLU}(\tilde{\mathbf{A}}_1^*\mathbf{X}_1^S\theta_1)])
\end{aligned}
\tag{A1.5}
$$

A1.2 Simulation and Case Study Model Setup

A1.2.1 Simulation model setup

Pre-trained HGCN: This work builds 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 the 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: This work designs 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 the 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: This work designs 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 recommended number of training epochs is 1000.

Pre-trained NN: This work designs 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 recommended number of training epochs is 1000.

Retrained NN layers: This work designs 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 recommended number of training epochs is 15000.

Baseline NN: This work designs 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 recommended number of training epochs is 5000.

A1.3 Case Study Model Setup

Pre-trained HGCN: This work builds 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 the 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: This work designs 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 the 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: This work designs 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 recommended number of training epochs is 1000.

Pre-trained NN: This work designs the NN, which has an input layer with 10 neurons, a hidden layer with 4 neurons, and an output layer with 3 neurons. The recommended number of training epochs is 1000.

Retrained NN layers: This work designs 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 recommended number of training epochs is 3500.

Baseline NN: This work designs 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 recommended number of training epochs is 6000.

Appendix 2

Appendix of Chapter 5

A2.1 Proof of Theoretical Results

Proof of Proposition 4.1 (OT Gradient as Barycentric Alignment)

For a response bin B_k, the bin-wise OT loss is

$$\lambda_{\mathsf{OT},k} = \sum_i \sum_j \pi_k^{(ij)} \|z_T^{(i)} - z_S^{(j)}\|^2 + \epsilon \sum_{i,j} \mathsf{KL}\left(\pi_k \| a_k^{(i)} b_k^{(j)\top}\right)$$

Fixing the transport plan π_k, the derivative with respect to a target client's latent feature is

$$\begin{aligned}\nabla_{z_T^{(i)}} \lambda_{\mathsf{OT},k} &= 2 \sum_j \pi_k^{(ij)} \left(z_T^{(i)} - z_S^{(j)}\right) \\ &= 2 \left(\sum_j \pi_k^{(ij)} z_T^{(i)} - \sum_j \pi_k^{(ij)} z_S^{(j)} \right)\end{aligned}$$

This expression shows that $z_T^{(i)}$ is pulled toward the barycenter of the source client's latent feature weighted by π_k.

Proof of Lemma 4.1 (Bounded OT Gradients)

From Proposition 4.1, this work assumes $\|z_T^{(i)}\|, \|z_S^{(j)}\| \le G$, and $w_k^{(i)} = \sum_j \pi_k^{(ij)}$. Then

$$\begin{aligned}\|\nabla_{z_T^{(i)}} \lambda_{\mathsf{OT},k}\| &\le 2\big(w_k^{(i)} \|z_T^{(i)}\| + \sum_j \pi_k^{(ij)} \|z_S^{(j)}\|\big) \\ &\le 2(w_k^{(i)} G + w_k^{(i)} G) = 4 w_k^{(i)} G\end{aligned}$$

Since the row sums of π_k satisfy $w_k^{(i)} \le 1$, the bound simplifies to

$$\|\nabla_{z_T^{(i)}} \lambda_{\mathsf{OT},k}\| \le 4G$$

Moreover, the Hessian is

$$\nabla^2_{z_T^{(i)}} \lambda_{\mathsf{OT},k} = 2 w_k^{(i)} I \succeq 0$$

establishing convexity in each latent coordinate.

Derivation of the Hessian. Differentiating the gradient expression again with respect to

$z_T^{(i)}$ gives

$$\nabla^2_{z_T^{(i)}} \lambda_{\mathrm{OT},k} = \nabla_{z_T^{(i)}} \left(2w_k^{(i)} z_T^{(i)} - 2 \sum_j \pi_k^{(ij)} z_S^{(j)} \right)$$

The second term is constant in $z_T^{(i)}$, so its derivative is zero. Thus,

$$\nabla^2_{z_T^{(i)}} \lambda_{\mathrm{OT},k} = 2w_k^{(i)} I,$$

which is positive semidefinite since $w_k^{(i)} \geq 0$. Therefore, the OT loss is convex in each latent coordinate.

Proof of Proposition 4.2 (Predictive Consistency via Conditional OT)

Suppose for each response bin B_k, the conditional latent distributions satisfy

$$W_\gamma\big(P(z_T \mid y_T \in B_k),\ P(z_S \mid y_S \in B_k)\big) \leq \gamma.$$

By definition of the Wasserstein distance, there exists a coupling $\pi_k \in \Pi(P(z_T \mid y_T \in B_k), P(z_S \mid y_S \in B_k))$ such that

$$\mathbb{E}_{(z_T, z_S) \sim \pi_k}\big[\|z_T - z_S\|\big] \leq \gamma$$

Step 1. Predictive distributions Bayes' rule gives the predictive distribution conditioned on the target client's latent features:

$$P(y_T \mid z_T) = \frac{P(z_T \mid y_T) P(y_T)}{P(z_T)}$$

Therefore, the distance of $P(z_T \mid y_T)$ and $P(z_S \mid y_T)$ determines how similar the predictive distributions $P(y_T \mid z_T)$ and $P(y_S \mid z_S)$ are when $P(z_T) \approx P(z_S)$ and $P(y_T) \approx P(y_S)$, which is mentioned in the assumption and alignment.

Step 2. Prediction based on Kantorovich--Rubinstein duality This work set $\phi(z) = \mathbf{1}_{\{g(z) \in A\}}$ be the indicator of a predictive event $A \subseteq \mathcal{Y}$, where g is the regressor mapping from latent features to response space. This function is bounded 1-Lipschitz under normalization. Therefore,

$$|P(y_T \in A \mid z_T) - P(y_S \in A \mid z_S)| \leq W_1\big(P(z_T \mid y_T), P(z_S \mid y_S)\big)$$

This inequality is based on the Kantorovich–Rubinstein theorem, it shows that for any 1-Lipschitz function ϕ,

$$\left|\mathbb{E}_{z_T}[\phi(z_T)] - \mathbb{E}_{z_S}[\phi(z_S)]\right| \leq W_1\big(P(z_T \mid y_T), P(z_S \mid y_S)\big)$$

Step 3. Apply the conditional OT bound Within each response bin B_k, by assumption,

$$W_\gamma\big(P(z_T \mid y_T \in B_k), P(z_S \mid y_S \in B_k)\big) \leq \gamma$$

Thus, for any measurable set $A \subseteq \mathcal{Y}$,

$$\sup_{A \subseteq \mathcal{Y}} |P(y_T \in A \mid z_T) - P(y_S \in A \mid z_S)| \leq h(\gamma).$$

where $h(\gamma) = \gamma$ for Wasserstein-1.

A2.2 Simulation and Case Study Setup

This section introduces the FedCOT model architectures and training configurations for the simulation and case study experiments. The best model structure and hyperparameters are selected based on trial and error and cross-validation for both the simulation and the case study.

A2.2.1 Simulation setup

This work uses simple architectures in the simulation since the simulated features are low-dimensional and generated from Gaussian distributions. This allows us to isolate the role of feature heterogeneity and validate the theoretical assumptions. Table A2.1 demonstrates the settings of the benchmark and FedCOT. All models in the simulation experiments use Xavier initialization for the linear layers, where the weights are drawn from a uniform distribution scaled according to the number of input and output units, and the biases are initialized to zero. This choice helps maintain stable gradients at the beginning of training.

Table A2.1: Simulation Setup for Different Models. Note: SF: sufficient features, IF: incomplete features.

Method	Encoder	Regressor	Optimizer	Epochs	Hyperparameters
Local-only (SF)	5-4-1 (ReLU)	1-16-1 (Sigmoid)	lr=1e-3, wd=1e-5	3000	–
Local-only (IF)	3-4-1 (ReLU)	1-8-1 (Sigmoid)	lr=1e-3, wd=1e-3	3000	–
FedAvg	3-4-1 (ReLU)	1-8-1 (Sigmoid)	lr=1e-3	3000	Rounds=10
FedProx	3-4-1 (ReLU)	1-8-1 (Sigmoid)	lr=1e-3	3000	Rounds=10, Proximal term=0.05
Ditto	3-4-1 (ReLU)	1-8-1 (Sigmoid)	lr=1e-3	3000	Rounds=10, $\lambda = 0.01$
FedKL	3-4-1 (ReLU)	1-16-1 (Sigmoid)	lr=1e-3	5000	λ_β=0.5
FedMMD	3-4-1 (ReLU)	1-16-1 (Sigmoid)	lr=1e-3	5000	λ_β=0.5
FedCOT (Ours)	3-16-1 (ReLU)	1-16-1 (Sigmoid)	lr=1e-4	12000	λ_β=0.5, OT (K=5, ε=0.05)

A2.2.2 Case study setup

Our case study utilizes deeper networks, showing the physical complexity and natural noise in the fatigue life prediction. Table A2.2 shows the setting of the benchmark and FedCOT. The same Xavier initialization scheme is adopted in the case study experiments. Using this initialization ensures consistent scaling across layers and improves the stability of optimization on the real dataset. Here, it is noticed that this work applied a log transformation for the distance of the defect to the surface and fatigue life, which are common operations in the fatigue life prediction problem. All the data has also been normalized by min-max normalization.

Table A2.2: Case Study Setup for Different Models

Method	Encoder	Regressor	Optimizer	Epochs	Hyperparameters
Local-only (SF)	7-64-4 (ReLU + Dropout-0.2)	4-8-1 (ReLU)	lr=1e-3	5000	–
Local-only (IF)	4-64-4 (ReLU + Dropout-0.2)	4-8-1 (ReLU)	lr=1e-3	2000	–
FedAvg	4-64-4 (ReLU + Dropout-0.2)	4-8-1 (ReLU)	lr=1e-3	5000	Rounds=10
FedProx	4-64-4 (ReLU + Dropout-0.2)	4-8-1 (ReLU)	lr=1e-3	5000	Rounds=10, Proximal term=0.05
Ditto	4-64-4 (ReLU + Dropout-0.2)	4-8-1 (ReLU)	lr=1e-3	5000	Rounds=10, $\lambda = 0.01$
FedKL	4-64-4 (ReLU + Dropout-0.2)	4-8-1 (ReLU)	lr=1e-4	10000	λ_β=0.5
FedMMD	4-64-4 (ReLU + Dropout-0.2)	4-8-1 (ReLU)	lr=1e-4	10000	λ_β=0.5
FedCOT (Ours)	4-64-4 (ReLU + Dropout-0.2)	4-8-1 (ReLU)	lr=1e-4	10000	λ_β=0.5, OT (K=5, ε=0.05)