

MACHINE LEARNING-BASED NON-DESTRUCTIVE EVALUATION
FOR FATIGUE DAMAGE QUANTIFICATION AND PROGNOSIS IN
REMANUFACTURING

Draft of April 17, 2021 at 17:18

BY

MIN-HSIU HSU

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Mechanical Science and Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2021

Urbana, Illinois

Adviser:

Assistant Professor Chenhui Shao

Abstract

Non-destructive evaluation (NDE) and fatigue damage assessment are crucial for quality control of remanufacturing processes. However, critical challenges exist in the development of NDE techniques for used components: individual NDE technology is only sensitive to specific fatigue conditions; analytics methods are lacking for quantitatively measuring accumulated mechanical damage and conducting prognostics in an early fatigue stage. In this paper, we propose a machine learning-based NDE technology by combining the strengths of linear ultrasonic (LU) and nonlinear ultrasonic (NLU) testings to characterize material properties and flaws at various length scales. Besides, a remaining useful life (RUL) estimation framework with hierarchical classifiers and S-N curves for identifying fatigue damage levels and inferring residual fatigue life of recycled parts is developed. In addition, regression models for estimating residual stress and full width at half maximum (FWHM) of X-ray diffraction (XRD) peak with ultrasonic testings are investigated. The effectiveness of the proposed methods is demonstrated using life cycle fatigue testing data for 5052-H32 aluminum alloy. This research aims to provide a screening system for end-of-life (EoL) products and lead to increasing usage of secondary materials for remanufacturing and high-quality products.

Draft of April 17, 2021 at 17:18

To my parents, for their love and support.

Acknowledgments

Table of Contents

List of Tables	vii
List of Figures	viii
LIST OF ABBREVIATIONS	ix
Chapter 1 Introduction	1
Chapter 2 Literature Review	5
2.1 Fatigue damage assessment	5
2.2 NDE of fatigue damage	5
2.3 LU and NLU applications	6
2.4 ML-based NDE	6
2.5 RUL estimation	7
2.6 Residual stress measurement with ultrasound	8
Chapter 3 Experiment	9
3.1 Life cycle fatigue testing	9
3.2 Interrupted fatigue testing	10
3.3 Linear and nonlinear ultrasound measurements	11
3.4 X-ray diffraction measurement	12
Chapter 4 Machine Learning Model Development	14
4.1 Signal pre-processing	14
4.2 Feature generation	15
4.3 Feature selection	17
4.4 Model training and validation	18
4.5 Hyperparameter tuning	19
Chapter 5 Remaining Useful Life Prediction	20
5.1 Problem formulation	20
5.2 Design of classifiers	21
5.3 RUL estimation with an S-N curve	28
5.4 Discussion	32

Chapter 6 Residual Stress and FWHM Prediction	35
6.1 Problem formulation	35
6.2 Residual stress prediction	36
6.3 FWHM prediction	38
6.4 Discussion	38
Chapter 7 Conclusion and Future Work	42
7.1 Conclusion	42
7.2 Future works	43
References	45

List of Tables

3.1	Summary of the interrupted fatigue testing specimens	11
5.1	Summary of the RUL prediction dataset	22
5.2	Summary of classification algorithms and model performance .	27
5.3	Classification performance of hierarchical classifier	29
5.4	RUL estimation for specimen 3	33
5.5	RUL estimation for specimen 5	34
6.1	Summary of the residual stress prediction dataset	36
6.2	Summary of the FWHM prediction dataset	36
6.3	Summary of regression models for residual stress estimation and model performance	37
6.4	Summary of regression models for FWHM prediction and model performance	38

List of Figures

1.1	General idea of the proposed methodology	1
1.2	Capability of defect detection for LU and NLU testings	2
3.1	Life cycle fatigue testing setup	10
3.2	S-N curve for 5052-H32 aluminum alloy reproduce the plot through python	10
3.3	Experimental setup for LU and NLU measurements at Prof. Matlack's Lab	12
3.4	Examples of linear and nonlinear ultrasonic signals	12
3.5	Schematic of the measurement locations for LU and NLU measurements. (The unit of length is in mm)	12
4.1	Overview of the ML model development	14
4.2	Flowchart for feature selection and hyperparameter tuning re-draw the flowchart	18
5.1	RUL estimation framework refine this figure	20
5.2	Design of multi-class classifier system	22
5.3	Design of multi-output classifier system	24
5.4	Design of two-stage classifier system	25
5.5	Design of hierarchical classifier system	26
5.6	Comparison of F1-score by class and classifier designs	28
5.7	Confusion matrix of hierarchical classifier for cross-validation .	30
5.8	Illustration of RUL inference procedure refine the figure	31
6.1	Scatter plot of actual vs predicted residual stress by RFECV-SVM axis label for MAPE	37
6.2	Scatter plot of actual vs predicted FWHM	40
6.3	Probability density plots for selected LU and NLU features .	41

LIST OF ABBREVIATIONS

Chapter 1

Introduction

Remanufacturing has increasingly received attention because of its energy-saving, eco-friendly, and cost-efficient characteristics. Remanufactured products have been presented in automotive, aerospace, and industrial machinery industries. For example, in the automotive industry, companies have been recycling and remanufacturing components such as engine, transmission, car body, etc., as a long-term tradition. In short, remanufacturing is a process of returning a used product to at least original performance specification from the customers' perspective. [citation?](#) To achieve this, secondary materials screening for quality control in remanufacturing processes by estimating the quantities of interest, e.g., remaining useful life (RUL) and residual stress, in incoming recycled end-of-life (EoL) products becomes an essential step and is crucial in increasing the usage of recycled materials. The general idea of this research paper is illustrated in Figure 1.1.

In recycled components, material fatigue damage is universally presented and it is one of the most influential factors that determine the RUL of a used product. Material fatigue has resulted in many catastrophic accidents in the history and has been studied for many decades; however, the fatigue damage level is hard to be monitored in real-world environments due to the

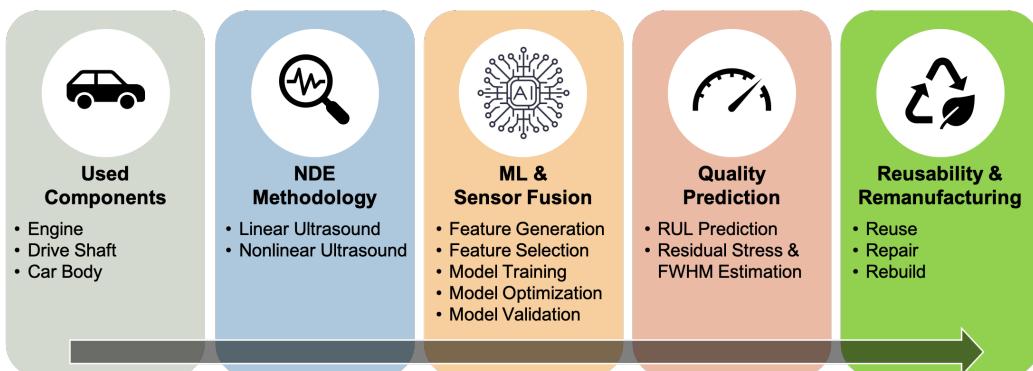


Figure 1.1: General idea of the proposed methodology

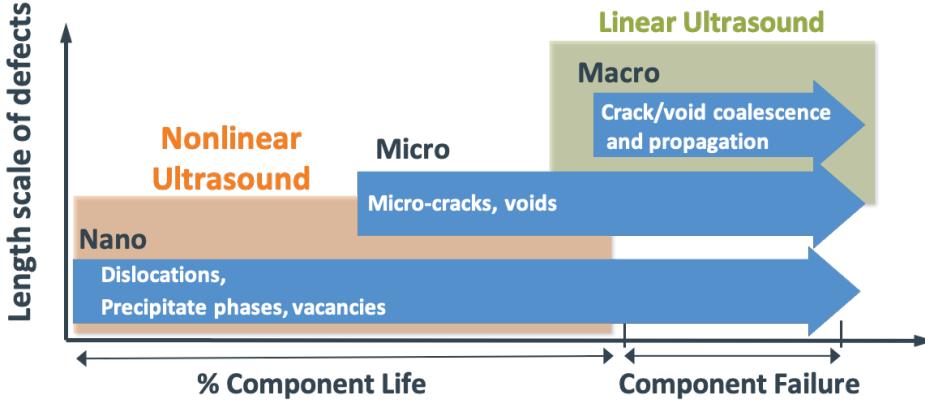


Figure 1.2: Capability of defect detection for LU and NLU testings

stochastic nature of fatigue behaviors and undetermined loading conditions [1], which is a critical issue to be addressed.

To quantitatively study fatigue damage in materials, non-destructive evaluation (NDE) methods have been developed [2]. NDE, also known as non-destructive testing (NDT), is a technique to evaluate material properties without causing damage to the testing parts. For instance, linear ultrasonic (LU) and nonlinear ultrasonic (NLU) testings send ultrasonic waves which propagate in a material and then analyze the response signals to evaluate material degradation. Although there exist a variety of NDE techniques, each of these methods is only sensitive to a few specific fatigue conditions and is limited to detecting defects in certain length scales. Figure 1.2 illustrates the detectable length scales of LU and NLU testings, where LU testings is robust at detecting macro-scale defects; in contrast, NLU techniques measure nonlinear material parameters to detect defects that are orders of magnitude smaller than the probing wavelength (in e.g. stainless steels, typically on the order of 1 mm), providing an important data source of early-stage damage characterization.

There is a lack of research on the estimation and prognosis of RUL of EoL components in the remanufacturing industry even though, with state-of-the-art machine learning (ML) models, RUL estimation has been successfully applied in many industrial components, e.g., bearing [3, 4, 5], gear [6], turbofan engine [7, 8, 9], and lithium-ion battery [7, 10, 11]. Unlike the examples in the literature, the possible difficulties for RUL estimation in remanufacturing are: (a) continuous and in-situ measurements are not available, i.e., unable

to know the historical measurements of a recycled component. (b) environmental noises can affect the performance of in-situ sensors as well as the built algorithms. (c) data in EoL products or remanufacturing is hard to collect and thus is not adequate to build robust data-driven models.

In this research, we proposed an ML-based NDE methodology for the quantification and prognostics of accumulated fatigue damage in recycled materials. First, we integrate LU and NLU testings to leverage the strength of each individual sensor, which has the potential of estimating different fatigue conditions simultaneously. Second, with multi-output hierarchical classifiers, an RUL estimation framework is developed by predicting the loading condition as well as the percentage of fatigue life that a component has undergone, and then the RUL is estimated from an S-N curve. In the proposed approach, both ultrasonic testings serve as ex-situ measurement methods to predict the loading history of a recycled component. Therefore, without continuous monitoring data of a component from its healthy state, the RUL can be inferred by using the current measurement only. Besides, a residual stress measurement method based on the proposed NDE methodology and ML techniques is also investigated, which shows the potential of being an efficient method in terms of test speed and cost.

This research is targeted at metals that are widely used in a number of industry sectors. In this preliminary study, the target material is 5052-H32 aluminum alloy which is widely used in truck and auto industries. Life cycle fatigue testings were conducted in different settings to construct a comprehensive database for fatigue progression in the target material. The fatigued specimens were later examined by ultrasonic as well as X-ray diffraction (XRD) measurements. Here, XRD measurements serve as the calibration for residual stress and full width at half maximum (FWHM) estimation. After that, the proposed NDE methodology and prediction framework were developed and tested on the collection of fatigued samples.

The goal of this research is to not only enable effective materials screening but also provide valuable information for the process optimization and control of downstream remanufacturing processes. As such, an effective NDE method that is applicable on the factory floor will lead to greater material recycling and improved quality of products that are produced using recycled materials.

The remainder of this paper is organized as follows. Chapter 2 provides the literature review of the related topics. Chapter 3 describes the experimental

setup and procedures including the design of experiments, life cycle fatigue testing, ultrasonic and XRD measurements. Chapter 4 introduces the ML model development procedure used in both Chapter 5 and Chapter 6 where the proposed RUL estimation framework and residual stress measurement method are presented, respectively, with the case study on our fatigue test dataset. Finally, in Chapter 7, the contributions of this thesis are summarized and future research directions are mentioned.

Chapter 2

Literature Review

2.1 Fatigue damage assessment

Fatigue damage is a critical issue in engineering due to the concern of safety, and accurate estimation of fatigue damage has been a decades-long study in areas such as remanufacturing, transportation equipment, and structural health monitoring. [Santecchia et al.](#) provides an extensive overview of fatigue damage models for metals from various perspectives including linear damage rule, continuum damage mechanics, multi-axial as well as variable amplitude loading, energy-based methods, and stochastic-based approaches. However, none of these can be universally accepted because of the complexity of fatigue damage behaviors in reality [1].

2.2 NDE of fatigue damage

In many practical scenarios, NDE has been adopted to quantify fatigue damage by investigating the correlation between measurement data and material deterioration [12]. Some common NDE techniques for evaluating fatigue damage are infrared thermography [13], holographic interferometry [14], microwave [15], ultrasonic testing [citation](#), magnetic methods [16], acoustic emission approaches [citation](#), and electrical resistance methods [citation](#). While numerous NDE methods are available, each of these techniques has its own characteristics and thus is only sensitive to only one or a few specific applications. Recently, [Wisner et al.](#) presented a review of NDE in material fatigue and stated that when combined, NDE methods have been shown to improve the robustness of damage detection by complementing each other [2] [order of citation](#).

2.3 LU and NLU applications

Among various NDE techniques, LU and NLU has demonstrated its applicability in fatigue damage assessment [17, 18, 19, 20], defect classification [21, 22], and residual stress measurement [23, 24, 25] in materials. In terms of fatigue damage, Joshi and Green showed that ultrasonic attenuation is an indicator of fatigue damage in experiments performed on aluminum and steel [17]. Nagy introduced an experiment setup to monitor the second-order acoustic-elastic coefficient during the cyclic loading test, and demonstrated that the change in a nonlinear parameter, which monotonically increases as a function of the number of cycles applied, is substantially more than the corresponding change in linear parameters, wave velocity and attenuation [18]. Matlack et al. presented a comprehensive review of second harmonic generation (SHG) measurements for the NDE of fatigue, thermal aging, and radiation-induced damage [19]. An analytical model developed by Cantrell used a material nonlinearity parameter β extracted from SHG measurements to quantify the level of dislocation substructures and cracks that evolve during cyclic fatigue of planar slip metals, presenting the potential of using SHG measurement to assess the remaining life of the material. However, practical implementation of this method requires that the loading and environmental conditions of fatigue are given [20].

2.4 ML-based NDE

In recent years, machine learning has been commonly applied to NDE techniques in the automated recognition of patterns in testing signals and the outcome of interest such as fatigue damage levels, defect types, and welding quality. For example, Baumgartl et al. implemented a CNN-based in-situ thermographic monitoring system to identify defects produced during the additive manufacturing process of H13 steel [26]. An optical interferometry-based real-time quality prediction system using ANN in laser beam welding was developed by Stadter et al. [27]. Loutas et al. proposed a framework utilizing AE signals for fatigue damage prognostics in composite materials with hidden semi Markov model and Bayesian neural network [28]. For ultrasonic testings, various features were engineered through fast Fourier trans-

form, wavelet transform, and statistical methods, and fed into ML models for defect classification in [21, 22]. ML-based NDE is fast growing and more applications such as RUL estimation and residual stress measurement exist in the literature.

2.5 RUL estimation

RUL estimation has received broad interests these days in many applications on industrial components such as bearing [3, 4, 5], gear [6], turbofan [7, 8, 9], and lithium-ion battery [7, 10, 11]. However, there are relatively fewer recent RUL researches directly studying material fatigue. In the early years, Ray and Tangirala presented a nonlinear stochastic model for predicting fatigue life for 2024-T3 aluminum alloy based on extended Kalman filter [29]. Peng et al. proposed a Bayesian updating framework based on a physics-based fatigue crack growth model with crack length estimated from piezoelectric sensor signals by a regressor to perform the lap joint fatigue life prognosis for 2024-T3 aluminum alloy [30]. Banerjee et al. recently utilized optical and acoustic NDE techniques accompanied with Kalman filter and particle filter to predict RUL in glass fiber reinforced polymer [31]. In the literature, most of the RUL research papers associated with fatigue damage rely on physical fatigue modeling and state-space models, e.g., particle filter method. Since physical models contain assumptions and approximation, physical models are limited in complex application scenarios.

Although lots of data-driven approaches for RUL estimation existed [32, 33] especially as deep learning has become increasingly prevalent these years [34], the successful cases generally require a sufficient amount of data for training the state-of-the-art models [4, 5, 8, 9, 11]. In the field of estimating residual fatigue life, Lim et al. developed a data-driven RUL prognosis technique with an artificial neural network (ANN) and nonlinear ultrasonic measurements for 6061-T6 aluminum [35]. There is, still, much fewer data-driven researches available for residual fatigue life estimation.

Moreover, we barely found methods that perform RUL prognostics by taking only the measurement at the current time step as input, which is the situation we envision for our application on materials screening processes in the remanufacturing industry. The existing examples using approaches such

as Kalman filter [29], particle filter [31], and recurrent neural network [9] make predictions based on successive measurements. To tackle the issue of lacking previous observations, one relevant research is [Mazhar et al.](#)'s work on EoL products, where the authors integrated Weibull analysis and an ANN model which takes a single measurement to assess the RUL of components for reuse. Due to the lack of data, nonetheless, synthetic data is needed for training the ANN in this work [36].

2.6 Residual stress measurement with ultrasound

Residual stresses are often existed in mechanical components and have been recognized as one of the main factors of fatigue failure [37]. Residual stress measurement, therefore, has been an active area. A review of recent progress of residual stress measurement from [GUO et al.](#) provides a comparison between a variety of methods from the aspects of resolution, applicable object, and limitations [38]. Specifically for ultrasonic testing, the theory of acoustoelastic effect (the presence of stress in solids causes changes in the speeds of ultrasonic waves) for measurement of residual stress has been studied in [23]. [Tanala et al.](#) compared ultrasonic velocity measurements with X-ray diffraction in determining residual stress across a steel pipe and an alloy plate, which stated that ultrasonic techniques are more efficient in test volume and the cost of equipment [24]. In recent studies, [Liu et al.](#) implemented a testing system to analyze the accuracy and feasibility of residual stress measurement in 6063-T4 aluminum alloy by ultrasonic longitudinal critically refracted wave based on acoustoelastic theory [25].

Chapter 3

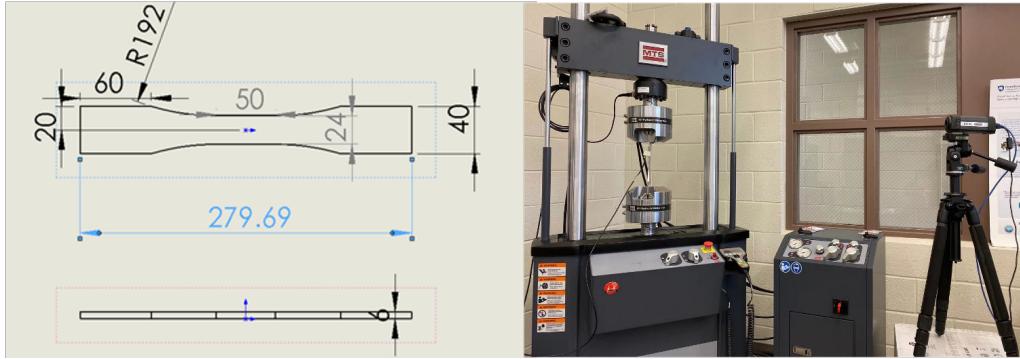
Experiment

This chapter describes the experimental procedure for collecting the life cycle fatigue test dataset used in this research for developing ML models. First, cyclic fatigue testings were conducted till the fracture of a specimen to acquire the fatigue characteristics of a material. Second, to mimic the scenarios in the remanufacturing industry, interrupted fatigue testing was utilized to produce specimens at different fatigue levels as a representation of end-of-life products. Then, LU and NLU measurements are used to evaluate the fatigue damage of those specimens stopped at the predetermined number of cycles in the interrupted fatigue test. Besides, the residual stress and (FWHM) data from XRD are also presented.

3.1 Life cycle fatigue testing

The life cycle fatigue testing aims to collect fatigue life data to understand the fatigue behavior of the target material. The fatigue life of a material is defined as the total number of cycles that a material can sustain under a specified loading condition. In order to develop the S-N curve of a material, the material is tested at different loading stress amplitudes, and the fatigue test is repeated multiple times for each loading stress amplitude to account for the variance of fatigue life.

The fatigue testing in this research is led by Prof. Li's group at the Penn State University. The target material is 5052-H32 aluminum alloy which is widely used for car body construction in the automotive industry. Figure 3.1 shows the dimension of the specimen and the test machine. Three loading amplitudes, 11.7, 12.7, and 14.7 kN for the cyclic fatigue testing are selected to develop the S-N curve which is shown in Figure 3.2.



(a) Schematic of the 5052-H32 aluminum alloy specimen **better schematic**

(b) MTS 100KN Landmark fatigue testing system at Prof. Jingjing li's lab

Figure 3.1: Life cycle fatigue testing setup

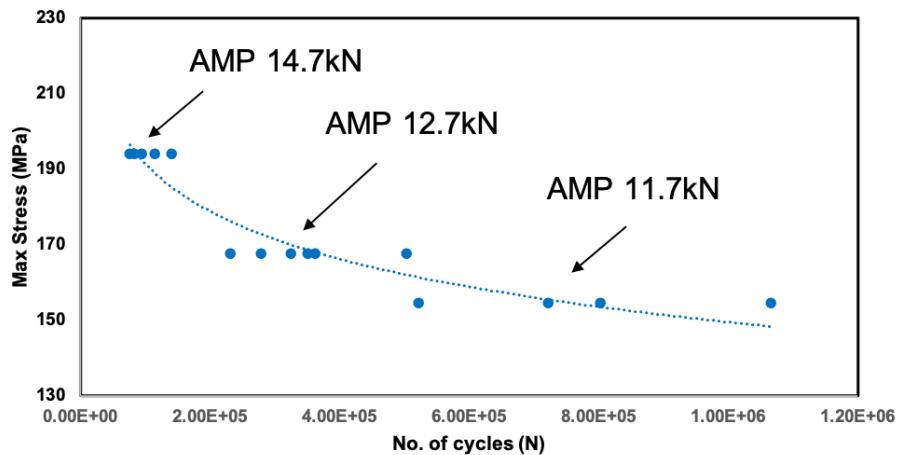


Figure 3.2: S-N curve for 5052-H32 aluminum alloy **reproduce the plot through python**

3.2 Interrupted fatigue testing

The purpose of performing interrupted fatigue testing is to produce damaged specimens at various fatigue levels by stopping the testing at several predetermined numbers of cycles. Considering the material cost and the time spent, the number of cycles applied to the specimens is set to be two levels, 33% and 67% fatigue life corresponding to the three loading amplitudes, 11.7, 12.7, and 14.7 kN. These specimens are used to represent the end-of-life products at different fatigue damage levels from the remanufacturing industry. Besides, three specimens without going through fatigue testing, i.e., 0%

Table 3.1: Summary of the interrupted fatigue testing specimens

Specimen ID	Loading Amplitude (kN)	Percentage of Fatigue Life (%)	Max Stress Applied (MPa)
1	11.7	33	176
2	11.7	33	176
3	11.7	67	176
4	11.7	67	176
5	12.7	33	195
6	12.7	33	195
7	12.7	67	195
8	12.7	67	195
9	14.7	33	221
10	14.7	33	221
11	14.7	67	221
12	14.7	67	221
13	—	0	—
14	—	0	—
15	—	0	—

fatigue life, are included as specimens at the healthy state. The summary of the interrupted fatigue testing specimens is presented in Table 3.1

3.3 Linear and nonlinear ultrasound measurements

In this research, the LU and NLU testings serve as the two main NDE methods for measuring the accumulated fatigue damage in the specimens. The ultrasonic testing is led by Prof. Matlack's group, and the testing system is shown in Figure 3.3. The LU and NLU measurements are both 1-D time-domain signals, but the two approaches differ based on different theories and parameters, e.g., excitation wave shape, frequency, and amplitude. Examples of LU and NLU signals are presented in Figure 3.4.

LU and NLU measurements were collected at nine locations on a specimen as illustrated in Figure 3.5, and each location was measured three times to ensure the measurement repeatability. As a result, for each specimen, there are $9 \times 3 = 27$ signal profiles produced. In total, the dataset contains 405 signal profiles.

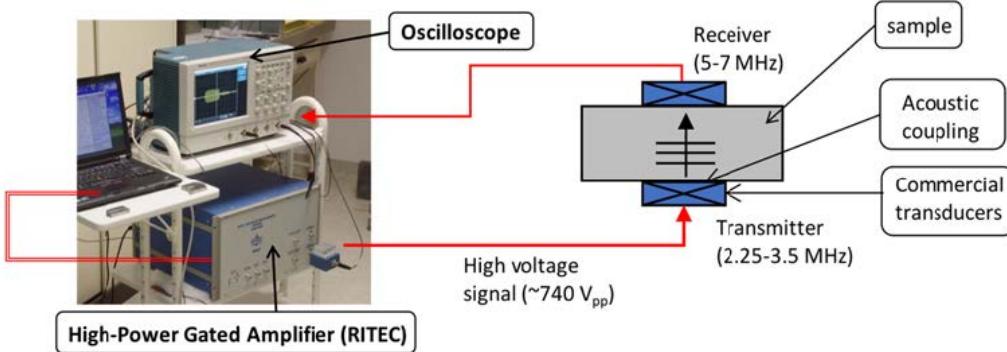
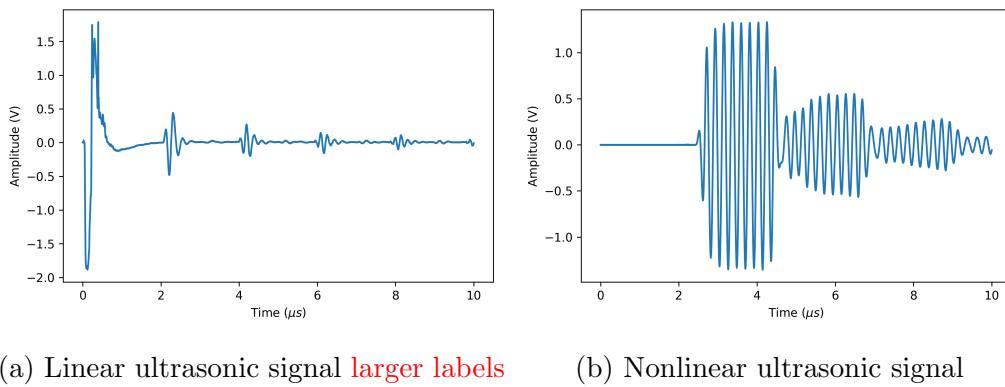


Figure 3.3: Experimental setup for LU and NLU measurements at Prof. Matlack's Lab



(a) Linear ultrasonic signal larger labels (b) Nonlinear ultrasonic signal

Figure 3.4: Examples of linear and nonlinear ultrasonic signals

3.4 X-ray diffraction measurement

Another quantity of interest, residual stress, is measured by XRD in this research. Residual stress is known to be associated with fatigue behaviors such as crack initiation and propagation. Besides, the FWHM height of the diffraction peak in XRD is also extracted. Prof. Li's group performed the XRD measurements for a subset of the specimens in the interrupted fatigue testing dataset. The residual stress and FWHM measured by XRD are used

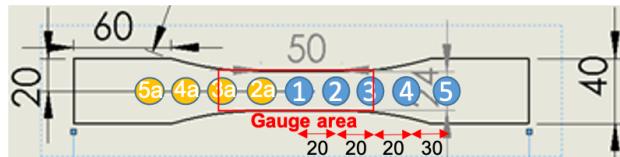


Figure 3.5: Schematic of the measurement locations for LU and NLU measurements. (The unit of length is in mm)

in the regression tasks in Chapter 6 as target variables.

Chapter 4

Machine Learning Model Development

This chapter introduces an ML model development procedure used in both classification tasks in Chapter 5 and regression tasks in Chapter 6. The procedure involves: (a) signal pre-processing, (b) feature generation, (c) feature selection, (d) model training, (e) model validation, and (f) hyperparameter tuning, as shown in Figure 4.1.

4.1 Signal pre-processing

It is essential to reduce noises and extract regions of interest in signals by signal processing before we perform analyses. **Figure XXX presents the signal pre-processing for LU and NLU signals.** First, DC bias is removed by subtracting the mean amplitude from a signal to prevent models from fitting on the bias. Second, considering the computational cost from the high-resolution data, we choose to downsample the ultrasonic signals. Third, we define the region of interest as the interval containing the response of an ultrasonic signal, and thus the other parts of a signal are discarded so that redundant information is not included.

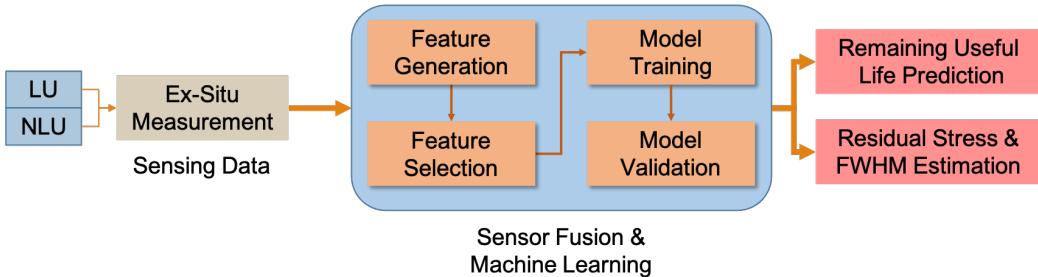


Figure 4.1: Overview of the ML model development

4.2 Feature generation

Since ultrasonic sensor signals are unstructured, which is difficult to be manipulated in an ML context, feature extraction methods are needed to create a representative set of values, i.e., features that aggregate the information from an entire signal. In this stage, physics-based and data-driven features are generated. The hybrid feature pool enables us to incorporate both physics knowledge and data-driven information into models.

4.2.1 Physics-based features

Given that physics modeling is built on theories or comprehensive experiment studies, physics-based features are robust, explainable, and suitable for applications having limited amounts of data such as the fatigue testing data in this research. Therefore, parameters (features) from traditional LU and NLU testings become potential candidates in the feature pool.

- Wave velocity

In LU testing, ultrasonic wave velocity is a stiffness-based measure that is associated with macroscopic damage such as crack/void coalescence and propagation. The wave speed is the distance divided by the time-of-flight that an ultrasonic wave propagating in a material, as shown by Equation (4.1)

$$v = \frac{2D}{\Delta t} \quad (4.1)$$

where wave velocity is denoted by v , and D is the thickness of a specimen. Δt is the time difference between the actuation pulse and the response signal. Notice that, in our LU testing setup, one transducer serves as both a transmitter and a receiver. Thus, the excitation signal travels $2D$ and the phase is changed 180° when received.

- Nonlinear acoustic parameter β

While wave velocity from LU testing is able to detect fatigue damage at the macroscale, it is limited because it cannot detect defects much smaller than the probing wavelength, e.g., 1 mm. In contrast, NLU

techniques are based on a different physical principle: nonlinear elasticity from nano- and micro-scale defects induce harmonic generation. The nonlinear acoustic parameter is related to the amplitude of generated harmonics. This nonlinear parameter changes due to defects such as dislocations, local plastic strain, precipitates, and micro-cracks, all of which are orders of magnitude smaller than the probing wavelength. Here, we apply fast Fourier transform (FFT) to NLU measurements and calculate the nonlinear parameter by using the ratio between the FFT amplitudes of the fundamental and the harmonic waves in Equation (4.2) **check the specific equation**

$$\beta = \frac{A_2}{A_1} \quad (4.2)$$

where A_1 , A_2 is the FFT amplitude of the fundamental wave and the second-order harmonic wave, respectively. **Figures for illustrating the calculations?**

4.2.2 Data-driven features

The physics-based features alone, however, are not enough to capture all of the information from the LU and NLU signals. As a result, a large number of features engineered from the time domain, frequency domain, and time-frequency domain of ultrasonic measurements are added to the feature pool.

- time domain features

The time domain features are peak amplitudes, ratios between peak amplitudes, and components from principal component analysis (PCA) and independent component analysis (ICA). Statistics in the time domain such as median, quantiles, variance, skewness, and kurtosis are also included. Besides, wave duration, wave energy, and the ratios between these quantities are calculated from the envelope analysis of an NLU signal.

- frequency domain features

Frequency domain analysis offers information that is not presented in the time domain. This information is especially valuable for periodic

signals such as ultrasonic measurements. Thus, after applying FFT, peak amplitudes, ratios between peak amplitudes, peak frequencies, frequency centroid, and frequency variance in FFT spectrum are extracted as the frequency domain features.

- Time-frequency domain features

Ultrasonic signals are usually not stationary, i.e., frequency changes in time, because of the interaction between ultrasonic waves and discontinuities within the material. Therefore, time-frequency analysis is needed to describe the phenomena. Discrete wavelet transform (DWT) is adopted to decompose ultrasonic measurements into several frequency bands [as shown in Figure XXX](#). Then, statistics such as mean, median, kurtosis, and skewness are recorded for each frequency band.

We concatenate features from LU and NLU signals together, and the feature pool contains 283 features in total. A list of candidate features for LU and NLU signal profiles is displayed in [Table XXX](#)

4.3 Feature selection

Feature selection aims to remove redundant features. Irrelevant features are common to see when we construct features without fully understanding a physical process such as the relationship between fatigue mechanism and ultrasonic responses. By including only the best subset of features for a prediction task, feature selection is beneficial for developing robust models against overfitting and improving model generalizability. There exist various feature selection techniques which can be mainly classified into three categories: filter methods, wrapper methods, and embedded methods. Each of these methods has its advantages, disadvantages, and applicable scenarios.

In the model development pipeline, we adopted a wrapper method, recursive feature elimination with cross-validation (RFECV) to obtain an optimal feature subset that achieves the best predictive performance in multiple training/test data splits for a single model. The RFECV algorithm is shown as a part of Figure 4.2. First, recursive feature elimination (RFE) starts from a set with all available features and eliminates k features step by step based on the feature ranking with regressors or classifiers until the predetermined

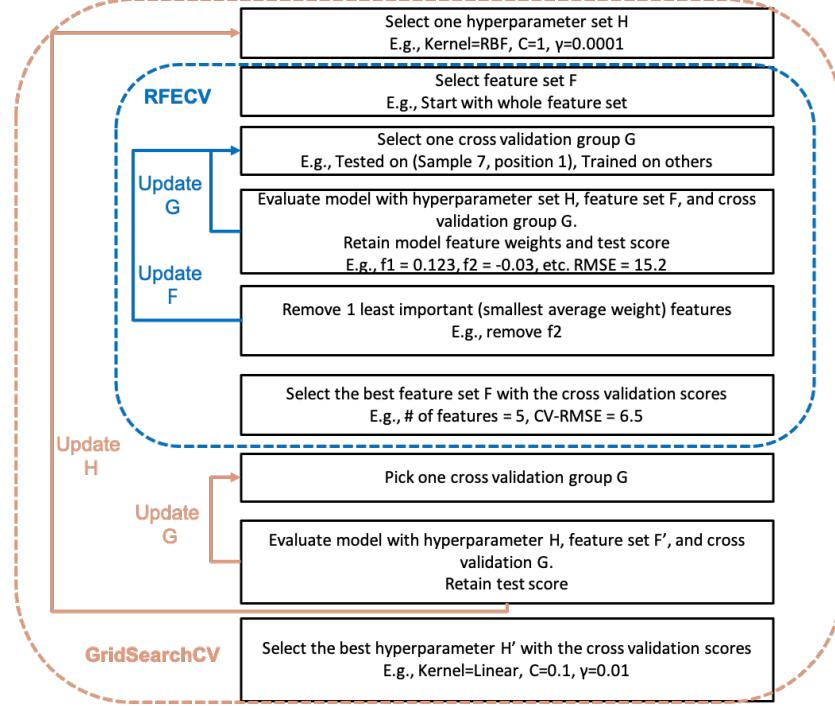


Figure 4.2: Flowchart for feature selection and hyperparameter tuning
re-draw the flowchart

number of features n is reached. Nevertheless, the best number of features to select n^* needs to be determined prior to the modeling. To find out n^* while alleviating the problem of overfitting, cross-validation (CV), a statistical model validation technique, is used along with RFE. CV partitions a dataset into a training set and a validation set in each iteration. A model is evaluated multiple times with different partitions, and n^* is determined by the overall validation results. Then, RFE selects the optimal n^* features from the feature pool. We choose a 5-fold CV in this feature selection procedure to avoid adding too much computational cost due to the fact that RFE is already computationally expensive.

4.4 Model training and validation

Model training and validation involve another CV loop in Figure 4.2 refine the flowchart to indicate this loop. However, this CV is not for finding the best feature subset but for providing a generalized estimate of a model's performance. Specifically, leave-one-group-out CV (LOGOCV) is applied,

where each group contains three repeated measurements at one measurement location on one specimen. Here, we make an assumption that each group, i.e., each measurement location on a specimen, is an independent sample because of the differences in the microstructure. In LOGOCV, each group is tested once and obtained validation scores by a predictor trained on the other groups, which efficiently utilizes the dataset and assesses the generalization capacity of a model.

4.5 Hyperparameter tuning

Searching for an optimal set of hyperparameters is another critical step that significantly influences ML model performance. We use grid search accompanied with the validation scores from the LOGOCV result to tune hyperparameters in a simple manner, as depicted in Figure 4.2. Grid search exhaustively considers all candidates from predefined hyperparameter combinations. Each combination is used to train a model, and the best hyperparameter set is the one achieving the best CV result.

The number of hyperparameters and the range of each hyperparameter vary in different learning algorithms. The candidate algorithms includes: logistic regression, support vector machine (SVM), and random forest for classification in Chapter 5; linear regression, Lasso regression, SVM, and random forest for regression in Chapter 6.

Chapter 5

Remaining Useful Life Prediction

In this chapter, we propose a framework for predicting the RUL of fatigue damaged samples based on ultrasonic testing. The framework has two parts: (a) an ML classification task and (b) an RUL inference procedure based on an S-N curve. First, ultrasonic signals are fed into ML classifiers to predict the loading condition and the percentage of fatigue life that a sample has gone through. Second, we estimate RUL from an S-N curve with the predicted loading condition and percentage of fatigue life. The proposed framework is visualized in Figure 5.1.

5.1 Problem formulation

Given the goal of predicting the RUL of EoL products, we formulate the task as an ML problem. In this section, We discuss possible formulations by considering the characteristics of the fatigue dataset and the impact on the ML system.

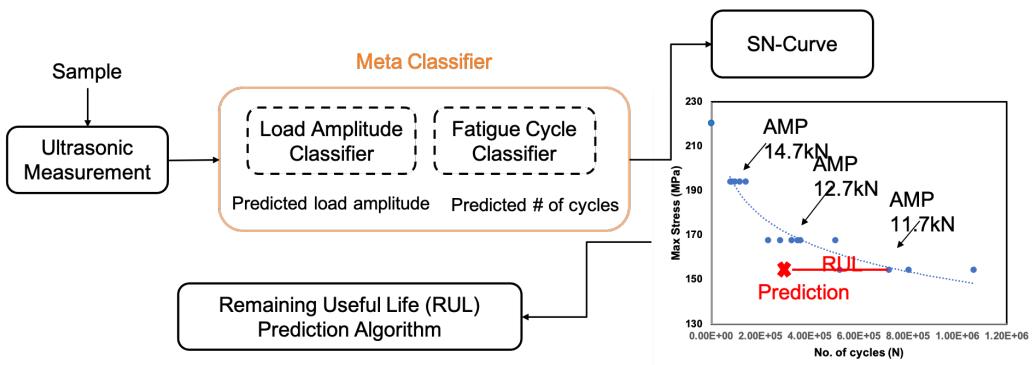


Figure 5.1: RUL estimation framework [refine this figure](#)

5.1.1 Dataset

In this RUL prediction task, the dataset is constructed on the ultrasonic measurements from the interrupted fatigue testing specimens in Table 3.1. There are 15 specimens and each of these was measured at 9 locations alongside 3 repeated measurements, producing 405 observations in total. Notice that we treat one measurement location on one specimen as a sample in the model training and validation procedure with LOGOCV, as described in Section 4.4. Besides, each specimen is tested by a combination of fatigue conditions from 4 loading amplitudes and 3 fatigue levels (the percentage of fatigue life), which forms the labels of a specimen.

5.1.2 Target variables

RUL is intuitively translated by the percentage of fatigue life that a sample has gone through. Normally, the percentage of fatigue life as a continuous target variable is treated as a regression task. However, since we only have 3 different percentages of fatigue life in the dataset, which is not ideal for regression modeling, we decide to view the percentage of fatigue life as a discrete variable and the problem becomes a classification task.

Loading amplitude is another target variable to be considered because loading condition affects the mechanism of fatigue damage in a material. For instance, at 33% fatigue life, a sample that experienced 11.7 kN loading and a sample that experienced 14.7 kN loading could exist different fatigue damages. Hence, we place a label that is a combination of loading amplitude and the percentage of fatigue life on each sample. Table 5.1 presents the labeled RUL prediction dataset.

5.2 Design of classifiers

In this section, classifiers are designed for predicting the loading condition (amplitude) and the percentage of fatigue life that a sample had experienced. Several classifiers are developed and the performance of each of those methods is evaluated based on the model development procedure in Chapter 4.

Table 5.1: Summary of the RUL prediction dataset

Specimen ID	Label (amplitude, percent fatigue life)	No. data
1		
2	Class 1 (11.7 kN, 33%)	54
3		
4	Class 2 (11.7 kN, 67%)	54
5		
6	Class 3 (12.7 kN, 33%)	54
7		
8	Class 4 (12.7 kN, 67%)	54
9		
10	Class 5 (14.7 kN, 33%)	54
11		
12	Class 6 (14.7 kN, 67%)	54
13		
14	Class 0 (0 kN, 0%)	81
15		

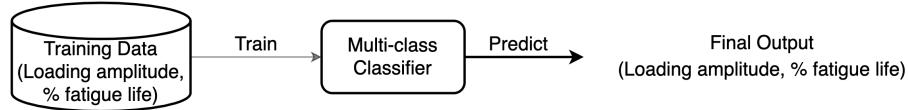


Figure 5.2: Design of multi-class classifier system

5.2.1 Multi-class classifier

A multi-class classifier is trained to classify a sample into 1 of the 7 classes. Figure 5.2 shows the training and inference process of the multi-class classifier. In multi-class classification problems, classes are mutually exclusive, e.g., class 1 and class 3 have nothing in common. However, despite we claimed that various loading conditions result in different fatigue behaviors in material, some similarities in material properties still exist. For example, at 33% fatigue life, samples with 11.7 kN loading and samples with 12.7 kN can have similar fatigue conditions. This idea can be applied to the commonality in samples at different percentages of fatigue life as well. With the assumption of mutual exclusivity, the multi-class formulation cannot capture these characteristics of the fatigue data.

5.2.2 Multi-output classifier

On the other hand, multi-output classification is capable of dealing with mutually non-exclusive classes by predicting the loading amplitude and percentage of fatigue life separately. A multi-output classifier outputs multiple labels, where predicting each label is considered a multi-class classification problem. In this design, we build one classifier for predicting the loading amplitude from input signals; another one for classifying a signal into one of the percentages of fatigue life. Then, the two predicted labels are combined based on a decision rule to output a class label that is consistent with the label in the dataset, as depicted in Figure 5.3.

These two classifiers, the loading amplitude classifier (LAC) and the fatigue cycle classifier (FCC), are trained separately with the same model development procedure but different target variables. As a result, unlike the single multi-class classifier trying to learn a way to separate 7 classes, the multi-output classification builds a LAC for a 4-class classification problem of the loading amplitude and an FCC for a 3-class problem of the percentage of fatigue life, which makes the problems easier to learn.

5.2.3 Two-stage classifier

Figure 5.4 shows the design of a two-stage classifier. Extended from the idea of the multi-output classifier in Subsection 5.2.2, a two-stage classifier is a classifier chain that predicts the loading amplitude and percentage of fatigue life in order. The classification starts with a LAC and the predicted loading amplitude is added to the feature space of an FCC, $\text{FCC}_{\text{two-stage}}$. By utilizing the predicted label as a feature for the next classifiers, the label dependence is preserved, i.e., the prediction of the percentage of fatigue life depends on the loading amplitude. In this case, we put the LAC before the $\text{FCC}_{\text{two-stage}}$ because sometimes the loading condition is given in real-life scenarios. Note that, in the training phase, the true loading amplitude is used to train the $\text{FCC}_{\text{two-stage}}$, but the $\text{FCC}_{\text{two-stage}}$ takes the predicted loading amplitude as one of the input features in inference.

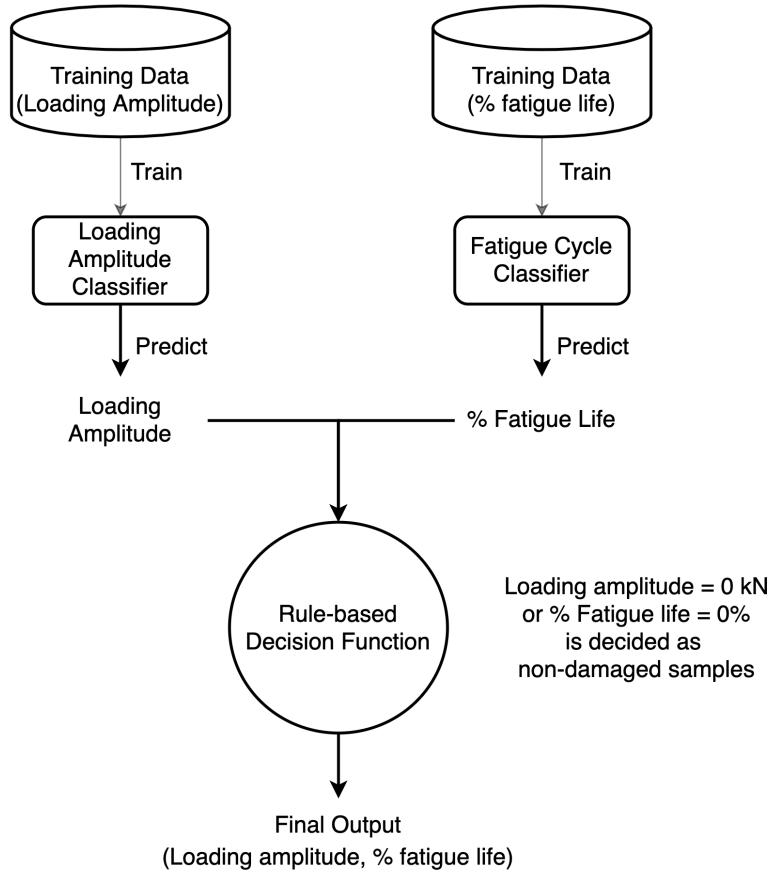


Figure 5.3: Design of multi-output classifier system

5.2.4 Hierarchical classifier

We further transform the multi-output task into a hierarchical classification scheme which is composed of multiple local classifiers based on a tree structure, shown in Figure 5.5. One advantage of the hierarchical classification is to exploit parent-child class relationships present in the class hierarchy. To achieve this, we train local classifiers per parent node in the taxonomy of the hierarchical classification problem. Specifically, one LAC and three FCCs for each loading amplitude are built; each of the FCCs is only trained on samples with the corresponding loading amplitude. For example, $FCC_{11.7\text{ kN}}$, where the subscript stands for the loading amplitude that the classifier corresponds to, is trained on samples with 11.7 kN loading applied.

A prediction is inferred by the following steps: (a) the LAC first outputs the loading amplitude from the LU and NLU input signals. (b) the pre-

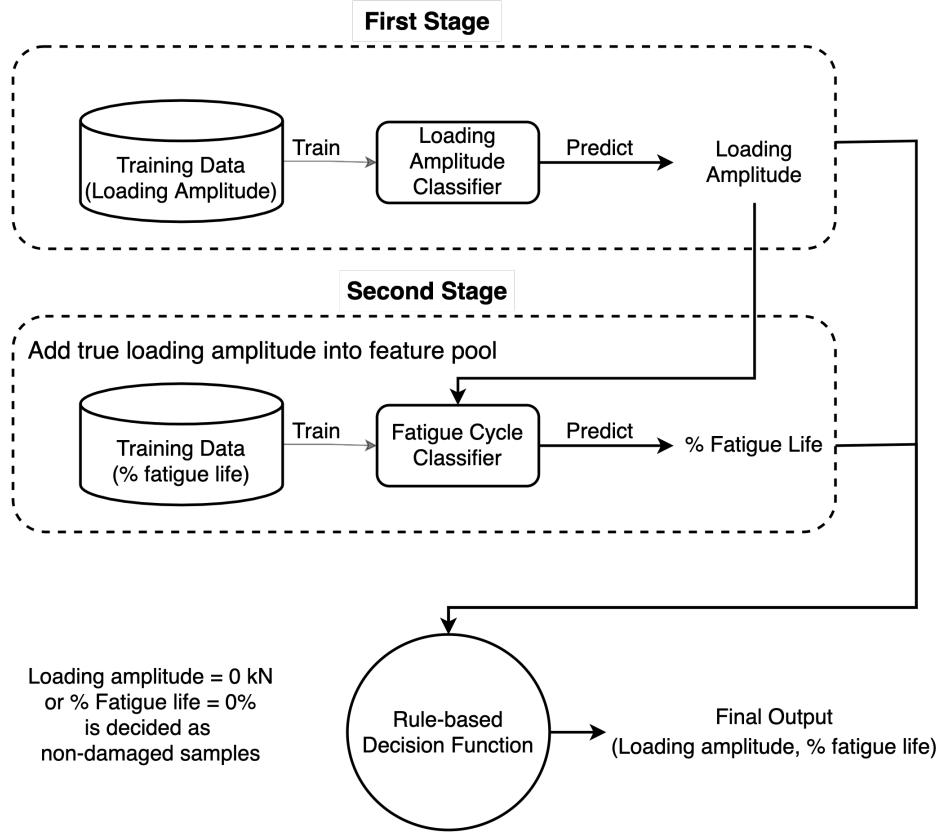


Figure 5.4: Design of two-stage classifier system

dicted loading amplitude is used to choose one of the FCCs for predicting the percentage of fatigue life. (c) the chosen FCC predicts the percentage of fatigue life. For example, 11.7 kN loading amplitude is predicted by the LAC. Therefore, $FCC_{11.7\text{ kN}}$ is selected and outputs a 33% fatigue life prediction. Finally, the predicted result is Class 1 (11.7 kN, 33%).

5.2.5 Evaluation metrics

Accuracy, recall, precision, and F1-measure are used to evaluate a classifier's performance with its LOGOCV result, and confusion matrices are also used to visualize model predictions. Although there exist other evaluation metrics for multi-output classification problems, we evaluate the aforementioned classifier designs in a unified multi-class classification problem with the label defined in Table 5.1. Thus, these classifier designs are comparable with each other. In this task, accuracy is an overall indicator for a model's performance since there is not much class imbalance in the dataset. Further-

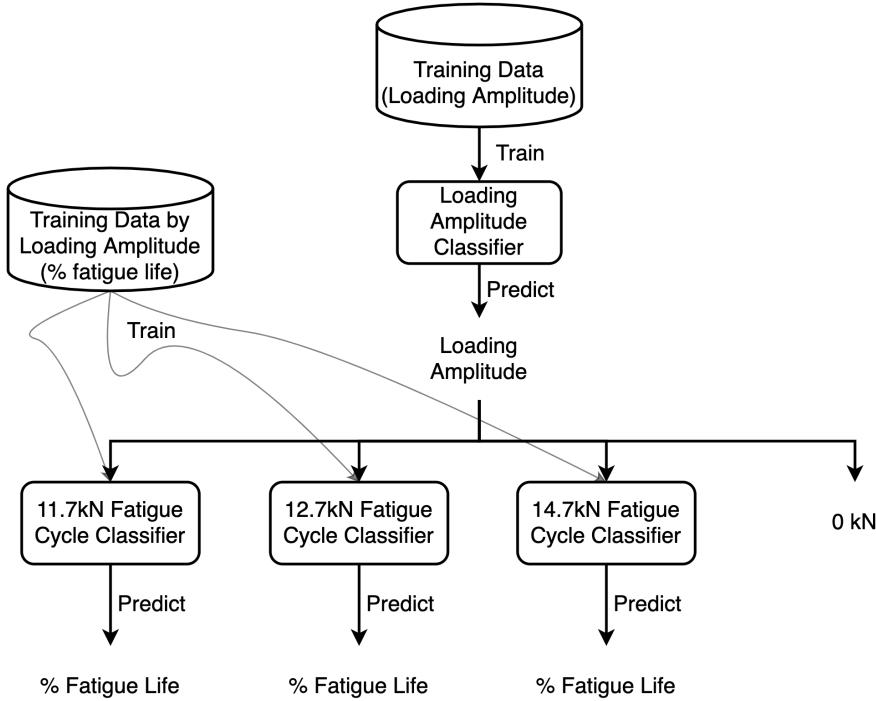


Figure 5.5: Design of hierarchical classifier system

more, a model's performance in each class is provided by recall, precision and F1-measure. Finally, confusion matrices serve as a visualization tool for summarizing the detailed result of the prediction from a classifier. The importance of each metric is left for practitioners to decide as it depends on different scenarios. For example, recall may be more important than precision in classifying highly-fatigued samples due to the consideration of safety.

5.2.6 Results

This subsection summarizes the results of the developed classifier designs. Table 5.2 presents the learning algorithms and the number of features determined by the model development procedure in Chapter 4. Figure 5.6 shows the performance of each classifier design by class. It is noticed that the proposed hierarchical classifier outperforms the other methods discussed in Section 5.2 for every class except class 6. Therefore, the hierarchical classifier is chosen to demonstrate the RUL estimation algorithm in the following sections.

The recall, precision, and F1-score of the hierarchical classifier are reported

Table 5.2: Summary of classification algorithms and model performance

Classifier Design	Learning Algorithm	No. Selected Features	LOGOCV Test Accuracy (%)
Multi-class	Logistic regression	35	77.3
Multi-output	Logistic regression	59 (LAC) 51 (FCC)	78.8
Two-stage	Logistic regression	59 (LAC) 51 (FCC _{two-stage})	81.5
Hierarchical	Logistic regression	59 (LAC) 10 (FCC _{11.7 kN}) 40 (FCC _{12.7 kN}) 6 (FCC _{14.7 kN})	85.3

in Table 5.3, and Figure 5.7 presents the detail of the LOGOCV classification result from the hierarchical classifier for 405 measurements in the dataset. It is observed that class 0 has 100% recall rate and only 2 measurements in class 5 are wrongly classified into class 0, implying that healthy samples are distinguishable from other damaged samples. Similarly, both class 1 and class 2 have 96.3% recall rate and few measurements are wrongly predicted as these two classes, showing that the classifier can identify the samples that had undergone the low-amplitude fatigue testing. In addition, even though classes 3-6 have more misclassified data points, those errors are mostly situated in the same loading amplitude as the true class, indicating the LAC is able to estimate the loading amplitude that a sample has been applied. With the overall accuracy being 85.2%, the proposed hierarchical classifier achieves a promising result for this application.

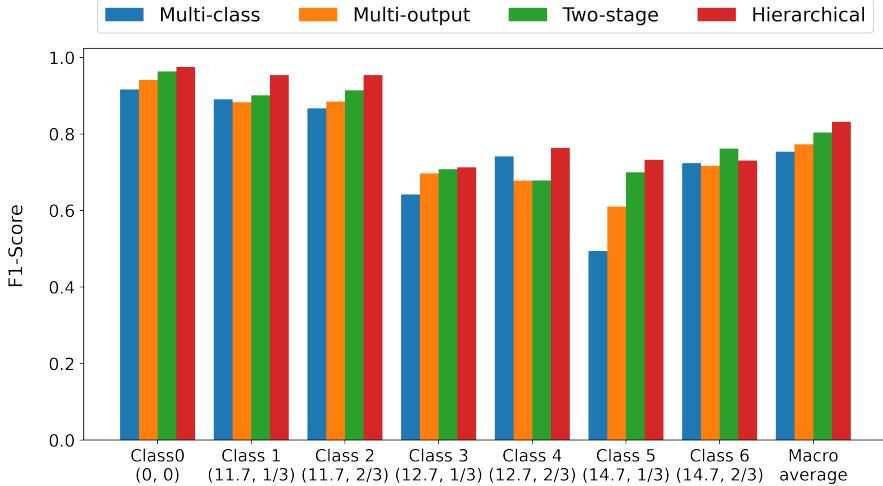


Figure 5.6: Comparison of F1-score by class and classifier designs

5.3 RUL estimation with an S-N curve

In the proposed framework, after obtaining the predicted loading amplitude and percentage of fatigue life, the RUL of a sample is estimated with the S-N curve of the material.

5.3.1 S-N curve with statistical distributions

Although an S-N curve is often referred to as the best-fit line of fatigue data, fatigue life data can be modeled as statistical distributions to account for the variability in fatigue life. The randomness in fatigue life comes from stochastic behaviors in material fatigue and the variance of microstructure in materials. As a result, the fatigue life is usually modeled by statistical distributions including Gaussian (normal), log-normal, and Weibull distribution [39]. Due to the limited amount of the available fatigue life data in this research, we use a Gaussian normal distribution to model the fatigue life X_a at a loading amplitude a as

$$X_a \sim Normal(\hat{\mu}_a, \hat{\sigma}_a^2) \quad (5.1)$$

where $\hat{\mu}_a$ is the sample mean, $\hat{\sigma}_a$ is the sample standard deviation of the fatigue life at the loading amplitude a .

Table 5.3: Classification performance of hierarchical classifier

	LOGOCV			No. data
	Precision	Recall	F1-score	
Class 0 (0 kN, 0%)	0.98	1.00	0.99	81
Class 1 (11.7 kN, 33%)	0.96	0.96	0.96	54
Class 2 (11.7 kN, 67%)	0.95	0.96	0.95	54
Class 3 (12.7 kN, 33%)	0.71	0.76	0.73	54
Class 4 (12.7 kN, 67%)	0.74	0.80	0.77	54
Class 5 (14.7 kN, 33%)	0.77	0.69	0.73	54
Class 6 (14.7 kN, 67%)	0.80	0.72	0.76	54
Macro Average	0.84	0.84	0.84	405
Weighted Average	0.85	0.85	0.85	405

5.3.2 RUL inference procedure

An S-N curve serves as a look-up table for linking a classifier's predictions to RUL estimation, and the illustration is provided in Figure 5.8. Also, the inference procedure is detailed below:

1. Plot the predicted loading amplitude and percentage of fatigue life in the S-N curve figure. For example, the prediction, class 1 (11.7 kN, 33% fatigue life), is plotted as the red cross in Figure 5.8.
2. Determine fatigue life with the corresponding loading amplitude from the S-N curve. The fatigue life can be in the format of a single value such as the mean fatigue life or a statistical distribution described in Subsection 5.3.1.
3. If the fatigue life is represented by a single value, RUL can be calculated by Equation (5.2):

$$RUL = N_{fl} - N_c \quad (5.2)$$

where N_{fl} is the fatigue life from an S-N curve and N_c is the number of

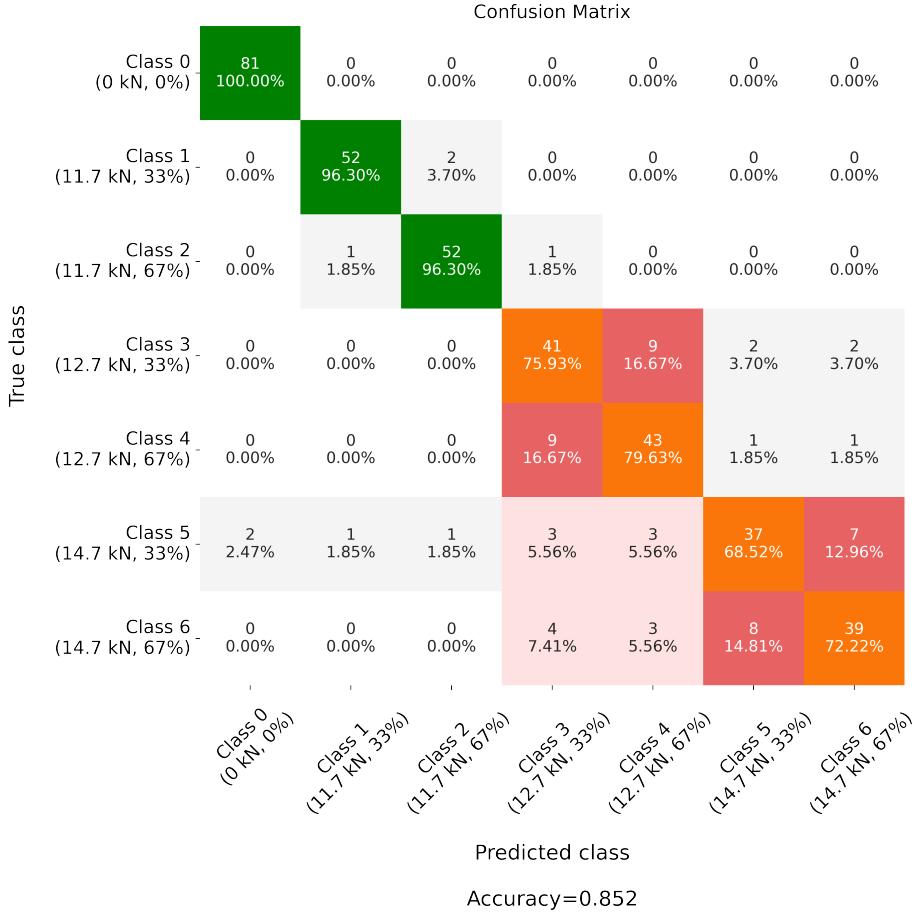


Figure 5.7: Confusion matrix of hierarchical classifier for cross-validation

cycles translated from the predicted percentage of fatigue life a sample has undergone.

If the fatigue life at a given loading amplitude is modeled as a normal distribution, the RUL of a sample can be estimated by transforming the fatigue life distribution by a factor of the predicted percentage of fatigue life, as in Equation (5.3) to (5.5):

$$X \sim Normal(\mu, \sigma^2) \quad (5.3)$$

$$Y = X - pX = (1 - p)X \quad (5.4)$$

$$Y \sim Normal((1 - p)\mu, (1 - p)^2\sigma^2) \quad (5.5)$$

where X follows a normal distribution $Normal(\mu, \sigma^2)$ with μ as the mean and σ^2 as the variance of fatigue life, and p is the predicted

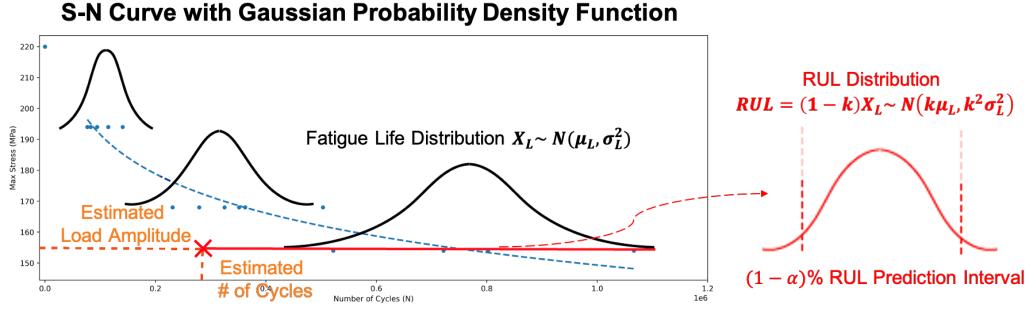


Figure 5.8: Illustration of RUL inference procedure [refine the figure](#)

percentage of fatigue life; Y is a random variable representing the RUL of a sample. Then, the point estimates, e.g., mean, median, or 25% quantile, as well as the interval estimate of RUL can be inferred. For instance, the 95% confidence interval of RUL is

$$CI = [(1 - p)\mu + z_{0.025}(1 - p)\sigma, (1 - p)\mu + z_{0.975}(1 - p)\sigma] \quad (5.6)$$

5.3.3 Examples

In this subsection, we demonstrate two examples of the RUL estimation based on the proposed framework. Because the true fatigue life of the testing specimens in the interrupted fatigue testing is not available, we cannot calculate the error of the RUL estimation for this stage. Instead, the 95% confidence intervals of the RUL of specimen 3 and 5 are presented in Table 5.4 and 5.5.

In Table 5.4, the 95% confidence intervals successfully cover a referencing RUL calculated by subtracting the recorded number of cycles from the mean fatigue life. Among all measurements for specimen 3, only 1 measurement at location 2a is misclassified, indicating the strength of the proposed model in estimating the RUL for samples that had undergone low-amplitude fatigue cycles. Also, the prediction of the other 2 repeated measurements at location 2a are correct, implying that an accurate decision can be made by considering multiple measurements at one location. For specimen 5 in Table 5.5, more measurements are misclassified. An interesting observation is, however, shorter RUL is estimated at location 1 and 2, at which higher damage levels

could happen since the two locations are closer to the center of the specimen.
Results for the other specimens are reported in Appendix XX.

5.4 Discussion

This section discusses several findings and limitations of the RUL estimation framework including (a) impacts from classifier designs and the structure of data, and (b) decision making based on multiple measurements. Other takeaways and issues are mentioned in Chapter 7.

5.4.1 Impacts from classifier designs and the structure of data

As presented in Section 5.2, a major problem can be decomposed into several sub-tasks by adopting a divide-and-conquer approach and the error rates can be reduced. In Table 5.2, one can observe that the prediction accuracy increases from 77.3% (multi-class classifier) to 85.3% (hierarchical classifier) as we split the meta classification problem into smaller classification tasks by exploiting the structure of the data. Nevertheless, the caveat is that introducing more classifiers for subtasks can increase the risk of overfitting the small dataset. The trade-off between the generalizability and the performance of a classifier design is left for further investigation.

5.4.2 Decision making based on multiple measurements

In the design of experiments, the ultrasonic measurements were conducted at different locations across the specimen and at one location repeatedly, offering opportunities for practitioners to make a decision based on the information from the spatial and repeated measurements. For example, one can determine the RUL of a specimen by selecting the minimum, average, or mode RUL among three repeated measurements at one location. Furthermore, the RUL of an entire specimen can be evaluated by similar or even more sophisticated approaches across multiple locations to decide the usability of a component. These rules are worth investigating in real application scenarios.

Table 5.4: RUL estimation for specimen 3

Specimen information

True label: Class 2 (11.7 kN, 67% fatigue life)

True number of cycles: 590193 cycles

Fatigue life from S-N curve: 777972 cycles

RUL calculated by true number of cycles: 187779 11.7 kN cycles

Replicate	Measurement Location				
	5a	4a	3a	2a	1
1	[132439, 386208]	[132439, 386208]	[132439, 386208]	[132439, 386208]	[132439, 386208]
	11.7 kN (2)	11.7 kN (2)	11.7 kN (2)	11.7 kN (2)	11.7 kN (2)
2	[132439, 386208]	[132439, 386208]	[132439, 386208]	[132439, 386208]	[132439, 386208]
	11.7 kN (2)	11.7 kN (2)	11.7 kN (2)	11.7 kN (2)	11.7 kN (2)
3	[132439, 386208]	[132439, 386208]	[264878, 772417]	[132439, 386208]	[132439, 386208]
	11.7 kN (2)	11.7 kN (2)	11.7 kN (1)	11.7 kN (2)	11.7 kN (2)

95% confidence interval in square brackets

Predicted class in parentheses

Bold text indicates misclassified measurements

Table 5.5: RUL estimation for specimen 5

Specimen information
 True label: Class 3 (12.7 kN, 33% fatigue life)
 True number of cycles: 90011 cycles
 Fatigue life from S-N curve: 342110 cycles
 RUL calculated by true number of cycles: 252099 12.7 kN cycles

Replicate	5a	4a	3a	2a	Measurement Location				
					1	2	3	4	5
1	[117963, 338183]	[117963, 338183]	[117963, 338183]	[117963, 338183]	[58981, 169091	[58981, 169091	[117963, 338183]	[117963, 338183]	[117963, 338183]
	12.7 kN (3)	12.7 kN (3)	12.7 kN (3)	12.7 kN (3)	12.7 kN (4)	12.7 kN (4)	12.7 kN (3)	12.7 kN (3)	12.7 kN (3)
	[58981, 169091	[117963, 338183]	[117963, 338183]	[117963, 338183]	[58981, 169091	[58981, 169091	[58981, 169091	[58981, 169091	[58981, 169091
2	[117963, 338183]	[117963, 338183]	[117963, 338183]	[117963, 338183]	12.7 kN (3)	12.7 kN (3)	12.7 kN (4)	12.7 kN (6)	12.7 kN (3)
	12.7 kN (4)	12.7 kN (3)	12.7 kN (3)	12.7 kN (3)	12.7 kN (4)	12.7 kN (4)	12.7 kN (6)	12.7 kN (3)	12.7 kN (3)
	[58981, 169091								
3	[117963, 338183]	[117963, 338183]	[117963, 338183]	[117963, 338183]	12.7 kN (3)	12.7 kN (3)	12.7 kN (4)	12.7 kN (3)	12.7 kN (3)
	12.7 kN (3)	12.7 kN (4)	12.7 kN (3)						
	[58981, 169091								

95% confidence interval in square brackets

Predicted class in parentheses

Bold text indicates misclassified measurements

Chapter 6

Residual Stress and FWHM Prediction

Because of the efficiency of ultrasonic testings in terms of inspection area and cost, we explore the potential of using ultrasonic testings to measure quantities of interest, residual stress and full width at half maximum (FWHM), which are originally obtained from XRD analysis. In this chapter, we present regression models for predicting the residual stress and the FWHM of XRD peaks in fatigue damaged samples based on the ultrasonic measurements.

6.1 Problem formulation

Following the same manner in Section 5.1, we first translate the prediction tasks into ML regression problems based on the available dataset.

6.1.1 Dataset

The dataset for predicting residual stress and FWHM is composed of the XRD results and ultrasonic measurements. To obtain the residual stress and FWHM, the XRD analysis was performed on a subset of samples, containing 8 specimens and 3 measurement locations for each specimen, in the RUL dataset. Table 6.1 and 6.2 are the summary of the residual stress and FWHM dataset, respectively. It is observed that specimen 7's relatively low FWHM values could indicate that there are microcrack initiations, and the discussion about cracks in specimen 7 is presented in Section 6.4.

6.1.2 Target variables

Residual stress and FWHM are the target variables in this chapter. Residual stress is known to influence fatigue behaviors including crack initiation and

Table 6.1: Summary of the residual stress prediction dataset

Specimen ID	Residual Stress (MPa)		
	Location 1	Location 2	Location 3
2	-61.7	-75.6	-80.2
4	-59.9	-69.6	-76.6
6	-60.3	-75.3	-79.6
7	-50.8	-59.6	-66.2
8	-57.3	-65.5	-79.7
10	-43.3	-47.0	-50.8
12	-38.8	-43.2	-50.0
14	-79	-76.7	-85.7

The negative sign indicates the compressive residual stresses.

Table 6.2: Summary of the FWHM prediction dataset

Specimen ID	FWHM ($^{\circ}$)		
	Location 1	Location 2	Location 3
2	0.354	0.353	0.355
4	0.350	0.354	0.353
6	0.358	0.359	0.363
7	0.307	0.320	0.321
8	0.357	0.355	0.358
10	0.356	0.358	0.360
12	0.354	0.353	0.355
14	0.338	0.340	0.346

propagation. FWHM is also an indicator for evaluating crack propagation. As a result, accurately predicting residual stress and FWHM based on ultrasonic measurements is beneficial to assist fatigue level estimation. Here, the problem is formulated as two regression tasks separately: (a) a univariate regression with residual stress as the target variable, and (b) a univariate regression with FWHM as the target variable.

6.2 Residual stress prediction

A regression model, RFECV-SVM_{RS}, for predicting residual stress based on ultrasonic signals is developed by following the procedure in Chapter 4. Besides, the RFECV-SVM_{RS} model is compared with other approaches such as Lasso regression, linear regression with top 5 features from Lasso regression, and random forest. Root mean squared error (RMSE) and mean

Table 6.3: Summary of regression models for residual stress estimation and model performance

Method	No. Selected Features	LOGOCV Test	
		RMSE (MPa)	MAPE (%)
Lasso regression	37	5.90	8.71
Linear regression	5	4.92	7.54
Random forest	283	7.74	12.85
RFECV-SVM	29	3.24	4.73

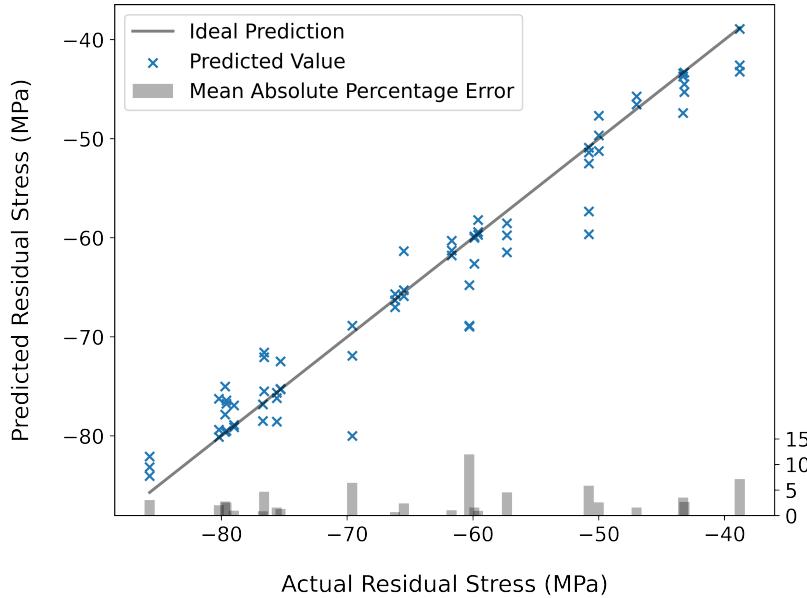


Figure 6.1: Scatter plot of actual vs predicted residual stress by RFECV-SVM axis label for MAPE

absolute percentage error (MAPE) are used to assess the model performance with LOGOCV results, as shown in Table 6.3. In this task, the RFECV-SVM_{RS} model performs the best with the MAPE equal 4.73%. Figure 6.1 illustrates the RFECV-SVM_{RS} model prediction by showing the actual and predicted residual stresses, in which a perfect model should follow the black line. It is observed that in most groups, at least one prediction among three repeated measurements (three predictions for the same actual residual stress in one sample group) is close to the ideal prediction, and the MAPEs for each group are mostly below 10%.

Table 6.4: Summary of regression models for FWHM prediction and model performance

Method	No. Selected Features	LOGOCV Test	
		RMSE (unit)	MAPE (%)
Lasso regression	20	0.0081	2.40
Linear regression	5	0.0056	1.62
Random forest	283	0.0099	2.81
RFECV-SVM	7	0.0063	1.89

6.3 FWHM prediction

Similarly, an FWHM prediction model, RFECV-SVM_{FWHM}, is built based on the procedure in Chapter 4 and is evaluated with the same metrics in Section 6.2. Table 6.4 is the comparison of the RFECV-SVM_{FWHM} with other regression methods. It is worth mentioning that the linear regression model with the top 5 features selected from Lasso regression achieves a 1.89% MAPE, performing slightly better than the RFECV-SVM_{FWHM}. Figure 6.2a and 6.2b illustrate the predictions of the linear regression model and the RFECV-SVM_{FWHM}, respectively. For the RFECV-SVM_{FWHM}, one can observe that the predictions of the small FWHM values from specimen 7 significantly deviate from the ideal predictions; nevertheless, the linear regression model is able to make close predictions. As mentioned earlier, we surmise that specimen 7 had developed cracks and the difference in the model performance led us to have a further investigation about specimen 7 in Section 6.4. Before performing further analysis, by excluding specimen 7, we can conclude that both models achieve a good performance in predicting FWHM with errors less than 2%.

6.4 Discussion

In this section, since we conjecture that specimen 7 had developed cracks based on the relatively small FWHM values measured from XRD analysis, the potential of using our ultrasonic technology to detect cracks is discussed.

6.4.1 Crack detection through LU and NLU measurements

To investigate whether our ultrasonic technology can detect cracks or not, we first assume specimen 7 is cracked and the differences in ultrasonic signals between specimen 7 and specimen 8 were resulted from the existence of cracks since specimen 7 and 8 were under the same experimental setting. Secondly, independent t-test is performed to test if the mean of a feature is significantly different between two groups, specimen 7 (15 measurements from location 3a to 3) and specimen 8 (15 measurements from location 3a to 3). Finally, Figure 6.3 displays several specimen 7's and specimen 8's probability density plots of the selected features. The distributions are not identical between specimen 7 and 8, indicating that these features are possible to distinguish cracked samples from normal samples; however, given that we only have two specimens in this test and there exist other factors causing the differences in measurement signals, further research on crack detection through our ultrasonic technology is required.

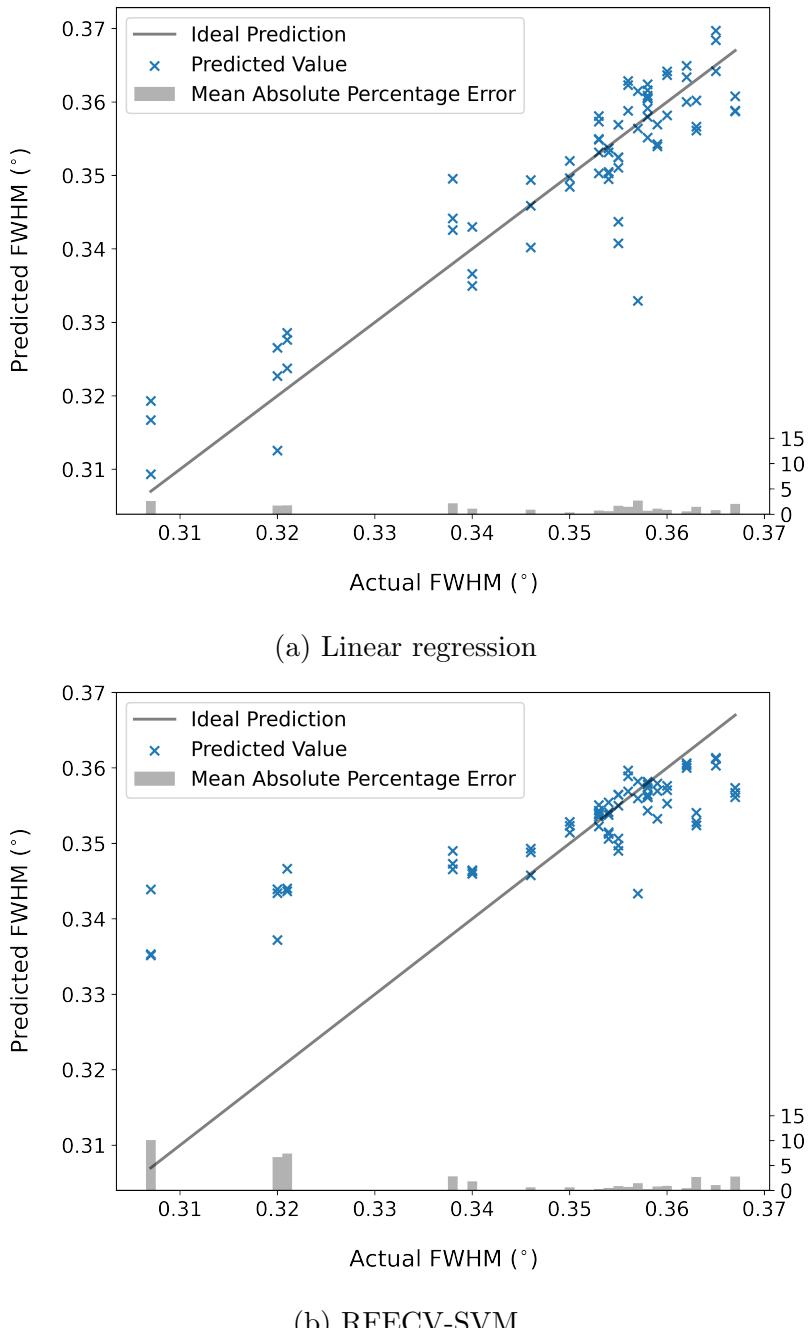


Figure 6.2: Scatter plot of actual vs predicted FWHM

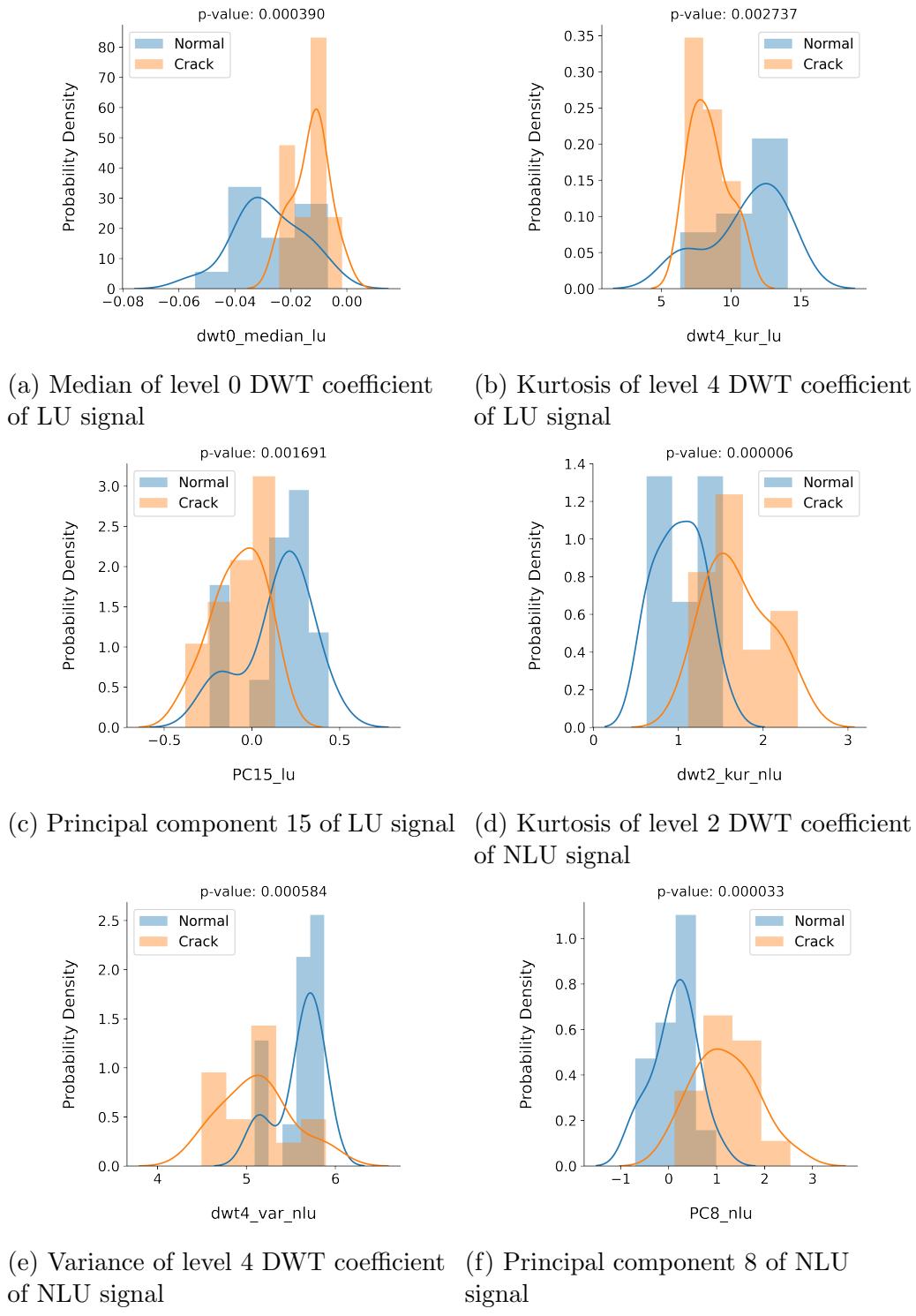


Figure 6.3: Probability density plots for selected LU and NLU features

Chapter 7

Conclusion and Future Work

7.1 Conclusion

An ML-based NDE methodology for assessing the accumulated fatigue damage level in terms of RUL, FWHM of XRD peak, and residual stress in recycled materials is developed, which aims to detect defects at various fatigue stages by combining the LU and NLU measurements and provide an ex-situ approach for the prognosis of useful life. An automatic pipeline is used to generate a pool of engineered features from raw ultrasonic signals, select useful features, validate models, and optimize classifiers and regressors.

A data-driven RUL estimation framework with the hierarchical classifiers and the statistical S-N curve is presented to bridge the research gap of RUL estimation in EoL products. The design of the hierarchical classification scheme utilizes the characteristics of the fatigue dataset to predict the 7 combinations of the loading amplitude and the percentage of fatigue life. Then, the use of statistical S-N curves incorporates the stochastic nature of fatigue life into the estimation of RUL. The framework relies on simple learning algorithms and does not require a large amount of data for training and validation, which shows the potential to be quickly adopted for other materials. More importantly, the framework does not need successive observations to conduct accurate prognostics of RUL; instead, one measurement at the time a sample is received is enough to provide RUL estimation, addressing the issue that the historical measurements may not be available for recycled components.

In addition, two regression tasks for predicting residual stress and FWHM using ultrasonic testings achieve high prediction accuracy. The little inconsistency between the ultrasonic predictions and the XRD measurements implies the potential to apply LU and NLU testings on measuring residual stress and

other fatigue damage indicators in a more cost-efficient and faster way.

We envision the proposed NDE methodology and the prediction framework will equip manufacturers with a responsive screening of incoming recycled materials, and lead to a significant increase in using recycled materials for remanufacturing and high-quality products that meet customer expectations.

7.2 Future works

In practice, nevertheless, there exist several limitations for the current work and suggested future research efforts are discussed in the following directions.

7.2.1 Collect more data for model development and validation

More data is needed to make our models generalizable in real-world applications even though we have utilized a cross-validation method, LOGOCV, to achieve good prediction accuracy while retaining the generalizability of the model performance with the limited amount of data. In the current LOGOCV, the group left out in the testing set is the three repeated measurements at one location; however, training data from neighboring locations on the same specimen still has high similarity with the testing group, which causes the potential data leakage in the training and validation phase. As a result, more data could enable a better estimate of model performance by treating each specimen as an individual group in LOGOCV.

True fatigue life, i.e., the number of cycles which a sample fails at, for an interrupted fatigue testing specimen is required to justify the proposed RUL estimation framework. Currently, the samples used to generate the predicted RUL were not tested until fracture, which prevents us from quantifying the error of our RUL prediction. Besides, more fatigue life data allows us to fit the distribution of fatigue life more precisely and/or to select a distribution family, e.g., Weibull, exponential, log-normal, etc., that better describes the fatigue life behavior. As a result, improvements in the accuracy and robustness of RUL estimation can be achieved.

Additionally, a finer measurement interval can extend the classification task so that classifiers can classify a sample into more fatigue levels, e.g., 15%, 33%, 50%, 67%, and 85%, or enable us to directly treat the RUL

prediction as a regression task. With this, more advanced models can be researched and better applicability as well as flexibility of the framework for practitioners are expected.

7.2.2 Multi-sensor fusion for fatigue damage assessment

The integration of multiple sensors has the capability of detecting different types of defects and improving the robustness of fatigue damage evaluation. However, this work only studies the combination of the LU and NLU testings. There are other NDE techniques such as infrared thermography, acoustic emission, etc., available to be added to the proposed NDE methodology. By integrating more sensors into the system, prediction performance can be further improved.

Moreover, sensor fusion can be categorized into several levels, e.g., data level, feature level, and decision level. We present a feature-level fusion of LU and NLU signals in this paper. Other fusion methods, especially decision-level fusion, are worth investigating because sensor selection can be conducted, by for example, discarding sensors that degrade the model performance at different fatigue stages. Therefore, a cost-effective quantitative evaluation of fatigue progression system can be designed.

7.2.3 Spatial and temporal modeling for fatigue damage evaluation

With the measurements being made at multiple spatial locations and at multiple time increments, spatiotemporal modeling/interpolation can be performed and provide a full map containing the fatigue damage information for a specimen in space and time. For example, spatial statistics can be used to relate the correlation between multiple measurement locations. Also, the temporal relation can be modeled if enough data measured at multiple life time of a component. Having the spatiotemporal map for fatigue evolution can give practitioners a better understanding of the decision-making process.

References

- [1] E. Santeccchia, A. M. S. Hamouda, F. Musharavati, E. Zalnezhad, M. Cabibbo, M. El Mehtedi, and S. Spigarelli, “A review on fatigue life prediction methods for metals,” *Advances in Materials Science and Engineering*, vol. 2016, p. 9573524, Sep 2016. [Online]. Available: <https://doi.org/10.1155/2016/9573524>
- [2] B. Wisner, K. Mazur, and A. Kortsos, “The use of nondestructive evaluation methods in fatigue: A review,” *Fatigue & Fracture of Engineering Materials & Structures*, vol. 43, no. 5, pp. 859–878, 2020. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/fme.13208>
- [3] J. Ben Ali, B. Chebel-Morello, L. Saidi, S. Malinowski, and F. Fnaiech, “Accurate bearing remaining useful life prediction based on weibull distribution and artificial neural network,” *Mechanical Systems and Signal Processing*, vol. 56-57, pp. 150–172, 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0888327014004087>
- [4] X. Li, Q. Ding, and J.-Q. Sun, “Remaining useful life estimation in prognostics using deep convolution neural networks,” *Reliability Engineering & System Safety*, vol. 172, pp. 1–11, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0951832017307779>
- [5] J. Shi, T. Yu, K. Goebel, and D. Wu, “Remaining Useful Life Prediction of Bearings Using Ensemble Learning: The Impact of Diversity in Base Learners and Features,” *Journal of Computing and Information Science in Engineering*, vol. 21, no. 2, 10 2020, 021004. [Online]. Available: <https://doi.org/10.1115/1.4048215>
- [6] B. He, L. Liu, and D. Zhang, “Digital Twin-Driven Remaining Useful Life Prediction for Gear Performance Degradation: A Review,” *Journal of Computing and Information Science in Engineering*, vol. 21, no. 3, 02 2021, 030801. [Online]. Available: <https://doi.org/10.1115/1.4049537>
- [7] A. Mosallam, K. Medjaher, and N. Zerhouni, “Data-driven prognostic method based on bayesian approaches for direct remaining useful life prediction,” *Journal of Intelligent Manufacturing*, vol. 27,

- no. 5, pp. 1037–1048, Oct 2016. [Online]. Available: <https://doi.org/10.1007/s10845-014-0933-4>
- [8] X. Li, Q. Ding, and J.-Q. Sun, “Remaining useful life estimation in prognostics using deep convolution neural networks,” *Reliability Engineering & System Safety*, vol. 172, pp. 1–11, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0951832017307779>
 - [9] J. Wu, K. Hu, Y. Cheng, H. Zhu, X. Shao, and Y. Wang, “Data-driven remaining useful life prediction via multiple sensor signals and deep long short-term memory neural network,” *ISA Transactions*, vol. 97, pp. 241–250, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0019057819302939>
 - [10] M. H. Lipu, M. Hannan, A. Hussain, M. Hoque, P. J. Ker, M. Saad, and A. Ayob, “A review of state of health and remaining useful life estimation methods for lithium-ion battery in electric vehicles: Challenges and recommendations,” *Journal of Cleaner Production*, vol. 205, pp. 115–133, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0959652618327793>
 - [11] K. Liu, Y. Shang, Q. Ouyang, and W. D. Widanage, “A data-driven approach with uncertainty quantification for predicting future capacities and remaining useful life of lithium-ion battery,” *IEEE Transactions on Industrial Electronics*, vol. 68, no. 4, pp. 3170–3180, 2021.
 - [12] J. Achenbach, “Quantitative nondestructive evaluation,” *International Journal of Solids and Structures*, vol. 37, no. 1, pp. 13–27, 2000. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0020768399000748>
 - [13] J. Fan, X. Guo, and C. Wu, “A new application of the infrared thermography for fatigue evaluation and damage assessment,” *International Journal of Fatigue*, vol. 44, pp. 1–7, 2012. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0142112312002046>
 - [14] T. Kreis, “Application of digital holography for nondestructive testing and metrology: A review,” *IEEE Transactions on Industrial Informatics*, vol. 12, no. 1, pp. 240–247, Feb 2016.
 - [15] A. J. K. M. Firdaus, R. Sloan, C. I. Duff, M. Wielgat, and J. F. Knowles, “Microwave nondestructive evaluation of thermal barrier coated turbine blades using correlation analysis,” in *2016 46th European Microwave Conference (EuMC)*, 2016, pp. 520–523.
 - [16] J. W. Hennek, A. Nemiroski, A. B. Subramaniam, D. K. Bwambok, D. Yang, D. V. Harburg, S. Tricard, A. K. Ellerbee,

- and G. M. Whitesides, “Using magnetic levitation for non-destructive quality control of plastic parts,” *Advanced Materials*, vol. 27, no. 9, pp. 1587–1592, 2015. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/adma.201405207>
- [17] N. R. Joshi and R. E. Green, “Ultrasonic detection of fatigue damage,” *Engineering Fracture Mechanics*, vol. 4, no. 3, pp. 577–583, 1972. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0013794472900677>
- [18] P. B. Nagy, “Fatigue damage assessment by nonlinear ultrasonic materials characterization,” *Ultrasonics*, vol. 36, no. 1, pp. 375–381, 1998, ultrasonics International 1997. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0041624X97000401>
- [19] K. H. Matlack, J.-Y. Kim, L. J. Jacobs, and J. Qu, “Review of second harmonic generation measurement techniques for material state determination in metals,” *Journal of Nondestructive Evaluation*, vol. 34, no. 1, p. 273, Nov 2014. [Online]. Available: <https://doi.org/10.1007/s10921-014-0273-5>
- [20] J. H. Cantrell, “Ultrasonic harmonic generation from fatigue-induced dislocation substructures in planar slip metals and assessment of remaining fatigue life,” *Journal of Applied Physics*, vol. 106, no. 9, p. 093516, 2009. [Online]. Available: <https://doi.org/10.1063/1.3254223>
- [21] S. Sambath, P. Nagaraj, and N. Selvakumar, “Automatic defect classification in ultrasonic ndt using artificial intelligence,” *Journal of Nondestructive Evaluation*, vol. 30, no. 1, pp. 20–28, Mar 2011. [Online]. Available: <https://doi.org/10.1007/s10921-010-0086-0>
- [22] G. Tripathi, H. Anowarul, K. Agarwal, and D. K. Prasad, “Classification of micro-damage in piezoelectric ceramics using machine learning of ultrasound signals,” *Sensors*, vol. 19, no. 19, 2019. [Online]. Available: <https://www.mdpi.com/1424-8220/19/19/4216>
- [23] C.-S. Man and W. Y. Lu, “Towards an acoustoelastic theory for measurement of residual stress,” *Journal of Elasticity*, vol. 17, no. 2, pp. 159–182, Jan 1987. [Online]. Available: <https://doi.org/10.1007/BF00043022>
- [24] E. Tanala, G. Bourse, M. Fremiot, and J. De Belleval, “Determination of near surface residual stresses on welded joints using ultrasonic methods,” *NDT & E International*, vol. 28, no. 2, pp. 83–88, 1995. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/096386959400013A>

- [25] H. Liu, Y. Li, T. Li, X. Zhang, Y. Liu, K. Liu, and Y. Wang, "Influence factors analysis and accuracy improvement for stress measurement using ultrasonic longitudinal critically refracted (lcr) wave," *Applied Acoustics*, vol. 141, pp. 178–187, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0003682X18302056>
- [26] H. Baumgartl, J. Tomas, R. Buettner, and M. Merkel, "A deep learning-based model for defect detection in laser-powder bed fusion using in-situ thermographic monitoring," *Progress in Additive Manufacturing*, vol. 5, no. 3, pp. 277–285, Sep 2020. [Online]. Available: <https://doi.org/10.1007/s40964-019-00108-3>
- [27] C. Stadter, M. Schmoeller, L. von Rhein, and M. F. Zaeh, "Real-time prediction of quality characteristics in laser beam welding using optical coherence tomography and machine learning," *Journal of Laser Applications*, vol. 32, no. 2, p. 022046, 2020. [Online]. Available: <https://doi.org/10.2351/7.0000077>
- [28] T. Loutas, N. Eleftheroglou, and D. Zarouchas, "A data-driven probabilistic framework towards the in-situ prognostics of fatigue life of composites based on acoustic emission data," *Composite Structures*, vol. 161, pp. 522–529, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0263822316316506>
- [29] A. Ray and S. Tangirala, "Stochastic modeling of fatigue crack dynamics for on-line failure prognostics," *IEEE Transactions on Control Systems Technology*, vol. 4, no. 4, pp. 443–451, 1996.
- [30] T. Peng, J. He, Y. Xiang, Y. Liu, A. Saxena, J. Celaya, and K. Goebel, "Probabilistic fatigue damage prognosis of lap joint using bayesian updating," *Journal of Intelligent Material Systems and Structures*, vol. 26, no. 8, pp. 965–979, 2015. [Online]. Available: <https://doi.org/10.1177/1045389X14538328>
- [31] P. Banerjee, R. P. Palanisamy, M. Haq, L. Udpa, and Y. Deng, "Data-driven prognosis of fatigue-induced delamination in composites using optical and acoustic nde methods," in *2019 IEEE International Conference on Prognostics and Health Management (ICPHM)*, June 2019, pp. 1–10.
- [32] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, "Remaining useful life estimation – a review on the statistical data driven approaches," *European Journal of Operational Research*, vol. 213, no. 1, pp. 1–14, 2011. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0377221710007903>

- [33] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, “Machinery health prognostics: A systematic review from data acquisition to rul prediction,” *Mechanical Systems and Signal Processing*, vol. 104, pp. 799–834, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0888327017305988>
- [34] S. Khan and T. Yairi, “A review on the application of deep learning in system health management,” *Mechanical Systems and Signal Processing*, vol. 107, pp. 241–265, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0888327017306064>
- [35] H. J. Lim, H. Sohn, and Y. Kim, “Data-driven fatigue crack quantification and prognosis using nonlinear ultrasonic modulation,” *Mechanical Systems and Signal Processing*, vol. 109, pp. 185–195, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S088832701830116X>
- [36] M. Mazhar, S. Kara, and H. Kaebernick, “Remaining life estimation of used components in consumer products: Life cycle data analysis by weibull and artificial neural networks,” *Journal of Operations Management*, vol. 25, no. 6, pp. 1184–1193, 2007, supply Chain Management in a Sustainable Environment Special Issue on Frontiers of Empirical Supply Chain Research. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0272696307000204>
- [37] G. Webster and A. Ezeilo, “Residual stress distributions and their influence on fatigue lifetimes,” *International Journal of Fatigue*, vol. 23, pp. 375–383, 2001. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0142112301001335>
- [38] J. GUO, H. FU, B. PAN, and R. KANG, “Recent progress of residual stress measurement methods: A review,” *Chinese Journal of Aeronautics*, vol. 34, no. 2, pp. 54–78, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1000936119304170>
- [39] H. Li, D. Wen, Z. Lu, Y. Wang, and F. Deng, “Identifying the probability distribution of fatigue life using the maximum entropy principle,” *Entropy*, vol. 18, no. 4, 2016. [Online]. Available: <https://www.mdpi.com/1099-4300/18/4/111>